Small Noise Asymptotics of the Bayesian Estimator in Nonidentifiable Models

Marc Joannides, François Le Gland

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Small Noise Asymptotics of the Bayesian Estimator in Nonidentifiable Models

Marc Joannides, François Le Gland

N ° 3675

Mai 1999

THÈME 4

Rapport de recherche

ISSN 0249-6399

Small Noise Asymptotics of the Bayesian Estimator in Nonidentifiable Models

Marc Joannides*, François Le Gland**

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet SIGMA2 (σ^2)

Rapport de recherche n3675 — Mai 1999 — 24 pages

Abstract: We study the asymptotic behavior of the Bayesian estimator for a deterministic signal in additive Gaussian white noise, in the case where the set of minima of the Kullback–Leibler information is a submanifold of the parameter space. This problem includes as a special case the study of the asymptotic behavior of the nonlinear filter, when the state equation is noise–free, and when the limiting deterministic system is nonobservable. As the noise intensity goes to zero, the posterior probability distribution of the parameter asymptotically concentrates on the submanifold of minima of the Kullback–Leibler information. We give an explicit expression of the limit, and we study the rate of convergence. We apply these results to a practical example where nonidentifiability occurs.

Key-words: Bayesian estimator, nonidentifiable model, nonlinear filtering, nonobservable system, small noise asymptotics, Laplace method, target motion analysis.

(Résumé : tsvp)

This work was partially supported by the Commission of the European Communities, under the HCM project *Statistical Inference for Stochastic Processes*, project number CHRX–CT92–0078, and by the Army Research Office, under grant DAAH04–95–1–0164.

* LABSAD, Université Pierre Mendès-France, 38040 GRENOBLE Cédex 9, France. — `joanni@labsad.upmf-grenoble.fr`
** IRISA / INRIA, Campus de Beaulieu, 35042 RENNES Cédex, France. — `legland@irisa.fr`

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : 02 99 84 71 00 - International : +33 2 99 84 71 00
Télécopie : 02 99 84 71 71 - International : +33 2 99 84 71 71

Comportement asymptotique de l'estimateur bayésien dans les modèles non–identifiables

Résumé : On étudie le comportement asymptotique de l'estimateur bayésien pour un signal déterministe observé dans un bruit blanc gaussien additif, dans le cas où l'ensemble des minima de l'information de Kullback–Leibler est une sous–variété de l'espace des paramètres. Ce problème inclut comme cas particulier l'étude du comportement asymptotique du filtre non–linéaire, quand l'équation d'état est non–bruitée, et quand le système déterministe limite est non–observable. Quand l'intensité du bruit tend vers zéro, la distribution de probabilité a posteriori du paramètre se concentre sur la sous–variété des minima de l'information de Kullback–Leibler. On donne une expression explicite de la limite, et on étudie la vitesse de convergence associée. On applique ces résultats à un exemple pratique où le modèle est non–identifiable.

Mots-clé : estimateur bayésien, modèle non–identifiable, filtrage non–linéaire, système non–observable, asymptotique petit bruit, méthode de Laplace, azimétrie.

1 Introduction

Consider the classical model of a deterministic signal depending on an unknown parameter, observed in additive Gaussian white noise, as studied in Ibragimov and Khasminskii [8, Chapter III, Section 5]. The d–dimensional observation $\{Y_t\,,\, 0 \leq t \leq T\}$ satisfies

$$dY_t = m_t(\theta)\, dt + \varepsilon\, dW_t^\theta \ ,$$

where $\theta \in \Theta$, the parameter set Θ is a Borel subset of $\mathbb{R}^p$, and $\{W_t^\theta\,,\, 0 \leq t \leq T\}$ is a standard Wiener process. The problem is to estimate the unknown parameter θ, given the observations $\{Y_t\,,\, 0 \leq t \leq T\}$. We assume that for any $\theta \in \Theta$, the mapping $t \longmapsto m_t(\theta)$ is measurable and satisfy the *finite energy* condition

$$\int_0^T |m_t(\theta)|^2\, dt < \infty \ .$$

In addition, we assume that for a.e. $0 \leq t \leq T$, the mapping $\theta \longmapsto m_t(\theta)$ is continuously differentiable, and for any $\theta \in \Theta$, the $p \times p$ symmetric nonnegative matrix (Fisher information matrix)

$$I(\theta) = \int_0^T [\dot{m}_t(\theta)]^*\, \dot{m}_t(\theta)\, dt$$

can be defined. Here and throughout the paper, the dot denotes derivation w.r.t. the parameter.

For any $\varepsilon > 0$, let $\{\mathbf{P}_\theta^\varepsilon\,,\, \theta \in \Theta\}$ be the family of probability measures generated on the canonical space $C([0,T]\,;\,\mathbb{R}^d)$ by the process $\{Y_t\,,\, 0 \leq t \leq T\}$ for different values of the parameter θ. The likelihood function for the estimation of θ based on $\{Y_t\,,\, 0 \leq t \leq T\}$ is given by

$$L^\varepsilon(\theta) = \exp\{\frac{1}{\epsilon^2} \int_0^T [m_t(\theta)]^*\, dY_t - \frac{1}{2\,\epsilon^2} \int_0^T |m_t(\theta)|^2\, dt\} \ .$$

Using the Bayesian approach, we model the *a priori* information on the unknown parameter θ by the prior probability distribution μ, and for any $\varepsilon > 0$ we denote by $\mathbf{P}^\varepsilon$ the probability distribution on the product space $C([0,T]\,;\,\mathbb{R}^d) \times \Theta$, defined by

$$\mathbf{P}^\varepsilon[A \times (\theta \in B)\,] = \int_B \mathbf{P}_\theta^\varepsilon(A)\, \mu(d\theta) \ .$$

The *posterior* probability distribution μ^ε is then defined by the Bayes rule

$$\langle \mu^\varepsilon, \phi \rangle = \mathbf{E}^\varepsilon[\phi(\theta) \mid \mathcal{Y}] = \frac{\displaystyle\int_\Theta \phi(\theta)\, L^\varepsilon(\theta)\, \mu(d\theta)}{\displaystyle\int_\Theta L^\varepsilon(\theta)\, \mu(d\theta)} \ ,$$

for any test function ϕ. The Bayesian point estimator $\widehat{\theta}^\varepsilon$ associated with a quadratic loss function coincides with the conditional mean, i.e.

$$\widehat{\theta}^\varepsilon = \int_\Theta \theta\, \mu^\varepsilon(d\theta) \ .$$

We introduce the *contrast process* $\ell^\varepsilon(\theta) = -\varepsilon^2 \log L^\varepsilon(\theta)$ and we denote by α the *true* value of the parameter. Discarding additional terms that are independent of θ, the contrast process can be written

$$\ell^\varepsilon(\theta) = -\varepsilon \int_0^T [m_t(\theta) - m_t(\alpha)]^*\, dW_t^\alpha + \tfrac{1}{2} \int_0^T |m_t(\theta) - m_t(\alpha)|^2\, dt \ ,$$

and converges to the Kullback–Leibler information

$$K_\alpha(\theta) = \tfrac{1}{2} \int_0^T |m_t(\theta) - m_t(\alpha)|^2\, dt \ ,$$

in $\mathbf{P}_\alpha^\varepsilon$–probability, as $\varepsilon \downarrow 0$.

Under the usual *identifiability* assumption that the true value α of the parameter is the only minimum point of K_α, the Bayesian estimator is consistant, i.e.

$$\widehat{\theta}^\varepsilon \longrightarrow \alpha \qquad \text{and} \qquad \mu^\varepsilon \Longrightarrow \delta_\alpha \ ,$$

in $\mathbf{P}^\varepsilon_\alpha$–probability, as $\varepsilon \downarrow 0$, and is asymptotically normal, i.e.

$$\frac{1}{\varepsilon}\,[\widehat{\theta}^\varepsilon - \alpha] \Longrightarrow \mathcal{N}(0,[I(\alpha)]^{-1}) \ ,$$

as $\varepsilon \downarrow 0$, provided α is in the interior of the parameter space, and provided the Fisher information matrix $I(\alpha)$ is invertible, see Ibragimov and Khasminskii [8, Chapter III, Section 5].

However, there are some practical situations, see Section 6 below, where *nonidentifiablilty* occurs, i.e. where the set of points with minimum Kullback–Leibler information does not reduce to the true value

$$M_\alpha = \operatorname*{argmin}_{\theta \in \Theta} K_\alpha(\theta) = \{\theta \in \Theta \,:\, m_t(\theta) = m_t(\alpha) \;\; \text{for a.e. } 0 \leq t \leq T\} \neq \{\alpha\} \ .$$

In this case, the point estimator $\widehat{\theta}^\varepsilon$ is not relevant, and we are rather interested in the asymptotic behavior of the *posterior* probability distribution μ^ε as $\varepsilon \downarrow 0$. It is easy to show that — although the set M_α has zero Lebesgue measure in general, hence $\mu^\varepsilon(M_\alpha) = 0$ for any $\varepsilon > 0$ — the probability distribution μ^ε is asymptotically supported by M_α, in the following sense : for any $c > 0$

$$\mu^\varepsilon(K_\alpha(\theta) < c) \longrightarrow 1 \ ,$$

in $\mathbf{P}^\varepsilon_\alpha$–probability, as $\varepsilon \downarrow 0$, see Proposition 3.2 below. This is only a qualitative result, and the question naturally arises whether more precise results could be obtained. In this paper, we show that it is possible to go beyond this qualitative result, provided higher order terms such as the Fisher information matrix are used.

Similar problems have been studied in the literature : The characterization of limit sets for the MLE has been obtained for diffusion–type processes observed over an infinite time horizon by Borkar and Bagchi [2], and for partially observed diffusion processes with small noise by James and LeGland [10]. In the simpler case where the set M_α is finite (a submanifold of dimension zero), the precise asymptotic behavior of the MLE and of the Bayesian point estimator has been studied for diffusion–type processes with small noise by Kutoyants [14] and [13, Section 2.7], and for general filtered statistical models by Kutoyants and Vostrikova [15].

The paper is organized as follows :

- In Section 2, we relate the problem of Bayesian estimation for nonidentifiable models, with the problem of nonlinear filtering for nonobservable systems.

- In Section 3, we introduce Assumptions A and B, which are regularity assumptions, Assumption C under which the set M_α of points with minimum contrast (Kullback–Leibler information) is a submanifold of $\mathbb{R}^p$, and Assumption D under which the prior probability distribution μ is absolutely continuous w.r.t. the Lebesgue measure, with a continuous density p.

- In Section 4, we characterize the limit as $\varepsilon \downarrow 0$ of the probability distribution μ^ε, as the *random* probability distribution

$$\mu_\alpha(dy) = c_\alpha \, \frac{\exp\{\frac{1}{2}\,|\xi_\alpha(y)|^2\}}{\sqrt{\det I_\perp(y)}} \; p(y)\,\lambda_\alpha(dy) \ ,$$

  where c_α is a normalizing constant, λ_α denotes the canonical measure on M_α, and where for any $y \in M_\alpha$, the matrix $I_\perp(y)$ denotes the restriction of the Fisher information matrix $I(y)$ to the normal space $N_y M_\alpha$, and $\xi_\alpha(y)$ is a Gaussian r.v. taking values in $N_y M_\alpha$.

- In Section 5, we study the corresponding rate of convergence, and we characterize the limit as $\varepsilon \downarrow 0$ of the conditional probability distribution of the r.v. $\dfrac{1}{\varepsilon}\,[\theta - \pi(\theta)]$, where π denotes the projection on M_α, as a mixture of *random* Gaussian probability distributions on the normal spaces to M_α.

- In Section 6, we present an application of our results to target motion analysis (TMA).

- Detailed proofs are given in Appendices A, B and C.

These results have been announced in Joannides and LeGland [11, 12].

2 Nonlinear filters for asymptotically nonobservable systems

As a prototypical situation where nonidentifiability occurs, consider a nonlinear filtering problem with noise–free dynamics, where the m–dimensional unobserved process $\{x_t\,,\,0 \leq t \leq T\}$ evolves according to the ODE

$$\frac{dx_t}{dt} = b(x_t)\ ,$$

with unknown initial condition x, and the d–dimensional observations are corrupted by some small additive Gaussian white noise

$$dY_t = h(x_t)\,dt + \varepsilon\,dV_t = h[\phi_t(x)]\,dt + \varepsilon\,dV_t\ ,$$

where $\{\phi_t(\cdot)\,,\,t \geq 0\}$ denotes the flow of diffeomorphisms associated with the ODE. In the Bayesian approach, the unknown initial condition x is given the prior probability distribution μ, and the problem consists in computing the posterior probability distribution μ^ε of the initial condition x given the observations $\mathcal{Y} = \sigma(Y_t\,,\,0 \leq t \leq T\}$, or equivalently computing the conditional probability distribution $\mu_T^\varepsilon = \mu^\varepsilon \circ \phi_T^{-1}$ of the final state x_T given the past observations $\mathcal{Y}$.

If the following deterministic system

$$(\Sigma) \qquad \left\{ \begin{array}{rcl} \dfrac{dx_t}{dt} & = & b(x_t) \\[2ex] z_t & = & h(x_t) \end{array} \right.$$

obtained in the limit as $\varepsilon \downarrow 0$, is *observable* on the time interval $[0,T]$, in the sense that the mapping

$$x \mapsto \{h[\phi_t(x)]\,,\,0 \leq t \leq T\}$$

from $\mathbb{R}^m$ to $C([0,T]\,;\,\mathbb{R}^d)$ is injective, then

$$\mu^\varepsilon \Longrightarrow \delta_{x_0}$$

in probability, as $\varepsilon \downarrow 0$, where x_0 denotes the *true* initial condition. The results of this paper allow to describe the asymptotic behavior of the posterior probability distribution μ^ε as $\varepsilon \downarrow 0$, when the limiting deterministic system (Σ) is *nonobservable*, i.e. when

$$\begin{aligned} M_0 &= \operatorname*{argmin}_{x \in \mathbb{R}^m} \int_0^T |h[\phi_t(x)] - h[\phi_t(x_0)]\,|^2\,dt \\ &= \{x \in \mathbb{R}^m\ :\ h[\phi_t(x)] = h[\phi_t(x_0)]\ \text{ for a.e. } 0 \leq t \leq T\} \neq \{x_0\}\ . \end{aligned}$$

In this example, the notion of nonobservability of the limiting deterministic system is equivalent to the notion of nonidentifiability of the corresponding statistical problem. In this context, the Fisher information matrix is defined by

$$I(x) = \int_0^T [\,h'[\phi_t(x)]\,\phi_t'(x)\,]^*\ h'[\phi_t(x)]\,\phi_t'(x)\,dt$$

for any $x \in \mathbb{R}^m$, and has been introduced in James [9] as the *observability Grammian* for the limiting deterministic system (Σ).

A typical nonlinear filtering problem where asymptotic nonobservability occurs, is target motion analysis (TMA), or tracking with bearings only measurements, see Lévine and Marino [17]. This application will be studied in more details in Section 6.

3 Preliminaries and assumptions

We assume that the parameter set Θ is a *compact* subset of $\mathbb{R}^p$, with nonempty interior. We first introduce the following *regularity* assumptions :

Assumption A : For a.e. $0 \leq t \leq T$, the mapping $\theta \longmapsto m_t(\theta)$ is continuously differentiable, and

$$\sigma_1 = \sup_{\theta \in \Theta} \sup_{|u|=1} \{ \int_0^T |\dot{m}_t(\theta)\, u|^2 \, dt \}^{1/2} < \infty \ .$$

Assumption B : For a.e. $0 \leq t \leq T$, the mapping $\theta \longmapsto m_t(\theta)$ is twice continuously differentiable, and

$$\sigma_2 = \sup_{\theta \in \Theta} \sup_{|u|=1} \{ \int_0^T |\ddot{m}_t(\theta)(u,u)|^2 \, dt \}^{1/2} < \infty \ ,$$

and for any $\theta, \theta' \in \Theta$

$$\sup_{|u|=1} \{ \int_0^T |\, [\ddot{m}_t(\theta) - \ddot{m}_t(\theta')]\, (u,u)|^2 \, dt \}^{1/2} \leq c_3\, |\theta - \theta'| \ .$$

Remark 3.1 For any $\theta \in \Theta$

$$\sup_{|u|=1} \int_0^T |\dot{m}_t(\theta)\, u|^2 \, dt = \sup_{|u|=1} u^* \, I(\theta)\, u = \lambda_{\max}(\theta) \ ,$$

where $\lambda_{\max}(\theta)$ denotes the largest eigenvalue of the Fisher information matrix $I(\theta)$, hence

$$\sup_{\theta \in \Theta} \lambda_{\max}(\theta) < \infty \ ,$$

under Assumption A.

Under Assumption A, the Kullback–Leibler information K_α is continuous, hence

$$\sigma_0 = \sup_{\theta \in \Theta} \{ \int_0^T |m_t(\theta) - m_t(\alpha)|^2 \, dt \}^{1/2} < \infty \ ,$$

and the Gaussian random field

$$\{ \int_0^T [m_t(\theta) - m_t(\alpha)]^* \, dW_t^\alpha \, , \, \theta \in \Theta \} \ ,$$

has $\mathbf{P}_\alpha^\varepsilon$–a.s. continuous sample paths, by the Kolmogorov continuity criterion, hence the r.v.

$$S_0 = \sup_{\theta \in \Theta} | \int_0^T [m_t(\theta) - m_t(\alpha)]^* \, dW_t^\alpha \, |$$

is $\mathbf{P}_\alpha^\varepsilon$–a.s. finite for any $\varepsilon > 0$. Moreover

$$\sup_{\theta \in \Theta} \mathbf{E}_\alpha^\varepsilon \, | \int_0^T [m_t(\theta) - m_t(\alpha)]^* \, dW_t^\alpha \, |^2 = \sigma_0^2 < \infty \ , \tag{1}$$

for any $\varepsilon > 0$. Similarly, under Assumption B, the Gaussian random fields

$$\{ \int_0^T [\dot{m}_t(\theta)\, u]^* \, dW_t^\alpha \, , \, \theta \in \Theta, |u| = 1\} \qquad \text{and} \qquad \{ \int_0^T [\ddot{m}_t(\theta)\, (u,u)]^* \, dW_t^\alpha \, , \, \theta \in \Theta, |u| = 1\}$$

have $\mathbf{P}_\alpha^\varepsilon$–a.s. continuous sample paths, by the Kolmogorov continuity criterion, hence the r.v.'s

$$S_1 = \sup_{\theta \in \Theta} \sup_{|u|=1} | \int_0^T [\dot{m}_t(\theta)\, u]^* \, dW_t^\alpha \, | \qquad \text{and} \qquad S_2 = \sup_{\theta \in \Theta} \sup_{|u|=1} | \int_0^T [\ddot{m}_t(\theta)\, (u,u)]^* \, dW_t^\alpha \, |$$

are $\mathbf{P}_\alpha^\varepsilon$–a.s. finite for any $\varepsilon > 0$. Moreover

$$\sup_{\theta \in \Theta} \sup_{|u|=1} \mathbf{E}_\alpha^\varepsilon \, | \int_0^T [\dot{m}_t(\theta)\, u]^* \, dW_t^\alpha \, |^2 = \sigma_1^2 < \infty \ , \tag{2}$$

and

$$\sup_{\theta\in\Theta}\sup_{|u|=1}\mathbf{E}_\alpha^\varepsilon\,|\int_0^T [\ddot{m}_t(\theta)\,(u,u)]^*\,dW_t^\alpha\,|^2 = \sigma_2^2 < \infty\ , \tag{3}$$

for any $\varepsilon > 0$.

The first qualitative result states that the posterior probability distribution μ^ε is asymptotically supported by M_α. The proof is given in Appendix A.

Proposition 3.2 *Under Assumption A, and if the prior probability distribution μ charges any neighborhood of M_α, then for any $c > 0$*

$$\mu^\varepsilon(K_\alpha(\theta) < c) \longrightarrow 1\ ,$$

in $\mathbf{P}_\alpha^\varepsilon$–probability, as $\varepsilon \downarrow 0$.

Under Assumptions A and B, the Kullback–Leibler information K_α is twice continuously differentiable, and the Fisher information matrix I is continuous. Notice that for any $y \in M_\alpha$, the matrix $I(y)$ coincides with the Hessian matrix $\ddot{K}_\alpha(y)$ of the Kullback–Leibler information : Indeed, for any $\theta \in \Theta$, the gradient and the Hessian matrix of the Kullback–Leibler information are respectively the row vector

$$\dot{K}_\alpha(\theta) = \int_0^T [m_t(\theta) - m_t(\alpha)]^*\,\dot{m}_t(\theta)\,dt\ ,$$

and the symmetric matrix

$$\ddot{K}_\alpha(\theta) = \int_0^T [m_t(\theta) - m_t(\alpha)]^*\,\ddot{m}_t(\theta)\,dt + \int_0^T [\dot{m}_t(\theta)]^*\,\dot{m}_t(\theta)\,dt\ .$$

Clearly $\dot{K}_\alpha(y) = 0$ and $\ddot{K}_\alpha(y) = I(y)$ for any $y \in M_\alpha$.

By the local inversion theorem, if $I(y)$ has full rank p, then y is an isolated minimum point of the Kullback–Leibler information. The case where this happens for any $y \in M_\alpha$ has been considered in Kutoyants and Vostrikova [15]. In this paper, we address the case where the set M_α of minimum points of the Kullback–Leibler information is a submanifold, and we make hereafter the following *rank* assumption :

Assumption C : The set M_α is contained in the interior of the parameter set Θ, and for any $y \in M_\alpha$, the Fisher information matrix $I(y)$ has constant rank $k \leq p$.

By the rank theorem, Assumption C ensures that M_α is a $(p-k)$–dimensional submanifold of $\mathbb{R}^p$.

Remark 3.3 By compactness, the set M_α has a finite number of connected components (possibly only one). The more general case where the Fisher information matrix has constant rank on any connected component, i.e. where M_α is the union of disjoint submanifolds with possibly different dimensions, could also be considered. Following the discussion in Hwang [7, Section 3], the only components that would matter in the limit, are those components where the rank is minimum, which corresponds to the highest dimensional submanifolds.

For any $y \in M_\alpha$, we denote respectively by T_yM_α and N_yM_α the tangent and normal spaces to M_α at point y. We easily check that

$$T_yM_\alpha = \ker I(y) \qquad \text{and} \qquad N_yM_\alpha = \operatorname{Im} I(y)\ .$$

Indeed, for any tangent vector v to M_α at point y, let $\{\gamma(s)\,,\, s \in I\}$ be a smooth curve in M_α such that $\gamma(0) = y$ and $\dot{\gamma}(0) = v$. For any $s \in I$ and for a.e. $0 \leq t \leq T$, $m_t[\gamma(s)] = m_t(\alpha)$, hence $\dot{m}_t[\gamma(s)]\,\dot{\gamma}(s) = 0$ and in particular for $s = 0$, $\dot{m}_t(y)\,v = 0$. Since the latter holds for a.e. $0 \leq t \leq T$, we have $I(y)\,v = 0$ which implies $T_yM_\alpha \subseteq \ker I(y)$, and also $\operatorname{Im} I(y) \subseteq N_yM_\alpha$ since for any $u \in \mathbb{R}^p$, $[I(y)\,u]^*v = 0$. By Assumption C, $\operatorname{Im} I(y)$ is a k–dimensional linear subspace of $\mathbb{R}^p$, hence $N_yM_\alpha = \operatorname{Im} I(y)$, which in turn implies $T_yM_\alpha = \ker I(y)$. In addition, for any $y \in M_\alpha$, the restriction $I_\perp(y)$ of the linear mapping $I(y)$ to its range N_yM_α is invertible, since it has full rank k. For later reference, notice that the smallest eigenvalue $\lambda_{\min}(y)$ of $I_\perp(y)$ is positive for any $y \in M_\alpha$. The above discussion shows that the Gaussian random field

$$\zeta_\alpha(y) = \int_0^T [\dot{m}_t(y)]^*\,dW_t^\alpha\ , \tag{4}$$

indexed by $y \in M_\alpha$ takes its values in $N_y M_\alpha$, since $v^* \zeta_\alpha(y) = 0$ for any $v \in T_y M_\alpha$. Therefore, the Gaussian random fields

$$\xi_\alpha(y) = [I_\perp(y)]^{-1/2}\,\zeta_\alpha(y) \qquad \text{and} \qquad \chi_\alpha(y) = [I_\perp(y)]^{-1}\,\zeta_\alpha(y) \tag{5}$$

indexed by $y \in M_\alpha$ are well defined, and take their values in $N_y M_\alpha$. Notice that

$$\sup_{y\in M_\alpha} |\zeta_\alpha(y)| = \sup_{y\in M_\alpha} \sup_{|u|=1} |\int_0^T [\dot{m}_t(y)\,u]^*\,dW_t^\alpha\,| \leq S_1 \ . \tag{6}$$

The next result clarifies the relationship between two different collections of neighborhoods for the manifold M_α : the tubular neighborhoods $M_\alpha^r = \{\theta \in \Theta : d(\theta, M_\alpha) < r\}$, and the sub–level sets $\{\theta \in \Theta : K_\alpha(\theta) < c\}$ of the Kullback–Leibler information. The proof is given in Appendix B.

Lemma 3.4 *Under Assumptions A to C*

$$\lambda_{\min} = \inf_{y\in M_\alpha} \lambda_{\min}(y) > 0 \qquad \textit{and} \qquad \lambda_{\max} = \sup_{y\in M_\alpha} \lambda_{\max}(y) < \infty \ ,$$

and for $r > 0$ small enough

$$\inf_{\theta\in\Theta\setminus M_\alpha^r} K_\alpha(\theta) \geq \tfrac{1}{4}\,\lambda_{\min}\,r^2 \qquad \textit{and} \qquad \sup_{\theta\in M_\alpha^r} K_\alpha(\theta) \leq \lambda_{\max}\,r^2 \ .$$

Indeed, it follows from Lemma 3.4 that the following inclusions hold, for $r > 0$ small enough

$$\{\theta \in \Theta : K_\alpha(\theta) < \tfrac{1}{4}\,\lambda_{\min}\,r^2\} \subset M_\alpha^r \subset \{\theta \in \Theta : K_\alpha(\theta) \leq \lambda_{\max}\,r^2\} \ .$$

We introduce finally the following *absolute continuity* assumption :

Assumption D : The prior probability distribution μ is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^p$, with a continuous density p.

4 Convergence

Under Assumptions A, C and D, we introduce the following *random* probability distribution on M_α

$$\mu_\alpha(dy) = c_\alpha\,\frac{\exp\{\frac{1}{2}\,|\xi_\alpha(y)|^2\}}{\sqrt{\det I_\perp(y)}}\;p(y)\,\lambda_\alpha(dy) \ ,$$

where c_α is a normalizing constant, and λ_α denotes the *canonical* (or Lebesgue) measure on M_α, see Berger and Gostiaux [1, Proposition 6.6.1]. We recall at this point that the normal space $N_y M_\alpha$ is also equipped with a canonical Lebesgue measure $\lambda_{\alpha,y}$, as a k–dimensional linear subspace of $\mathbb{R}^p$.

Notice that both μ^ϵ and μ_α are random measures, i.e. r.v.'s with values in the set $\mathcal{P}(\mathbb{R}^p)$ of probability measure on $\mathbb{R}^p$. Since μ^ϵ is absolutely continuous w.r.t. the Lebesgue measure, and μ_α is supported by a set of zero Lebesgue measure, the topology associated with the total variation distance is obviously too strong, and we use instead the topology associated with the following norm on $\mathcal{P}(\mathbb{R}^p)$

$$\|\mu - \mu'\|_{\mathrm{BL}}^* = \sup_{\|\phi\|_{\mathrm{BL}}=1} |\,\langle\mu,\phi\rangle - \langle\mu',\phi\rangle\,| \ ,$$

where $\|\cdot\|_{\mathrm{BL}}$ is the norm on the space of bounded and Lipschitz continuous functions, i.e.

$$\|\phi\|_{\mathrm{BL}} = \max\{\|\phi\|, \|\phi\|_{\mathrm{L}}\} \ ,$$

where $\|\cdot\|$ and $\|\cdot\|_{\mathrm{L}}$ denote respectively the supremum and the Lipschitz norm. The dual norm $\|\cdot\|_{\mathrm{BL}}^*$ is equivalent to the Prokhorov distance, and defines the same topology as the topology associated with the weak convergence of probability measures, see Dudley [3, 4]. We now state the first result of this paper, a generalization of which is given in Theorem 5.3.

Theorem 4.1 *Under Assumptions A to D, and if the prior density p is positive on M_α, then*

$$\|\mu^\epsilon - \mu_\alpha\|^*_{\mathrm{BL}} \longrightarrow 0 \ ,$$

in $\mathbf{P}^\epsilon_\alpha$–probability as $\epsilon \downarrow 0$.

Remark 4.2 Precisely, the following exponential rate is achieved : for any $\eta > 0$

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\kappa(\varepsilon)} \ \log \mathbf{P}^\varepsilon_\alpha[\, \|\mu^\varepsilon - \mu_\alpha\|^*_{\mathrm{BL}} > \eta] < 0 \ ,$$

where $\kappa(\varepsilon) \sim \varepsilon^{-2/5} \ \delta(\varepsilon)$, and $\delta(\varepsilon)$ goes to zero arbitrarily slowly as $\varepsilon \downarrow 0$.

Remark 4.3 In principle, the standard Wiener process $\{W^\alpha_t \, , \, 0 \leq t \leq T\}$ depends on ε through the definition

$$dW^\alpha_t = \frac{1}{\varepsilon} \ [dY_t - m_t(\alpha) \, dt] \ ,$$

and consequently the limit probability distribution μ_α should also depend on ε. However, the result as stated in Theorem 4.1 makes perfect sense, and in addition the probability distribution of μ_α does not depend on ε.

Sketch of the Proof. From the definition of the norm $\| \cdot \|^*_{\mathrm{BL}}$, it is enough to look at the difference

$$\langle \bar{\mu}^\varepsilon, \phi \rangle - \langle \bar{\mu}_\alpha, \phi \rangle \ ,$$

between the unnormalized probability distributions $\bar{\mu}^\varepsilon$ and $\bar{\mu}_\alpha$ defined by

$$\langle \bar{\mu}^\varepsilon, \phi \rangle = \varepsilon^{-k} \int_\Theta \phi(\theta) \ \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(\theta)\} \, p(\theta) \, d\theta \ , \tag{7}$$

and

$$\langle \bar{\mu}_\alpha, \phi \rangle = (2\,\pi)^{k/2} \int_{M_\alpha} \phi(y) \ \frac{\exp\{\frac{1}{2} \, |\xi_\alpha(y)|^2\}}{\sqrt{\det I_\perp(y)}} \ p(y) \, \lambda_\alpha(dy) \ ,$$

respectively. The approach is to obtain suitable estimates for this difference, holding at least on some *good* sets, and to control the probability of the complementary *bad* sets.

The integral in (7) is viewed as a Laplace integral, hence for small values of ε, we expect its behavior to be determined only by the set of minimum points of ℓ^ε. As $\varepsilon \downarrow 0$, this set shrinks to M_α, so we are naturally lead to evaluating this integral over a small neighborhood of M_α. Following the same approach as in Hwang [7, Section 3], the first step consists in replacing the integral over the whole space Θ by the integral over the tubular neighborhood

$$M^r_\alpha = \{\theta \in \Theta \ : \ d(\theta, M_\alpha) < r\} \ ,$$

and to show that the contribution of $\Theta \setminus M^r_\alpha$ is exponentially small as $\varepsilon \downarrow 0$. The next step is to evaluate the integral over M^r_α using a Fubini–like theorem. Since M_α is compact, then by the tubular neighborhood theorem, see Berger and Gostiaux [1, Theorem 2.7.12], for $r > 0$ small enough

$$M^r_\alpha \simeq \{(y,u) \ : \ y \in M_\alpha, u \in N_y M_\alpha, |u| < r\} = N^r M_\alpha \ ,$$

under the diffeomorphism

$$j_\alpha \ : \ (y,u) \in N^r M_\alpha \longmapsto \theta = y + u \in M^r_\alpha \ .$$

It is proved in Weyl [19] that the Jacobian determinant J_α of the transformation j_α is a positive continuous function on $N^r M_\alpha$, identically equal to 1 on M_α, i.e. for $u = 0$. The change of variable $\theta = y + u$ in the integral over M^r_α yields

$$\begin{aligned} &\varepsilon^{-k} \int_{M_\alpha} \{ \int_{N_y M^r_\alpha} \phi(y+u) \ \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(y+u)\} \, p(y+u) \, J_\alpha(y,u) \, \lambda_{\alpha,y}(du) \ \} \, \lambda_\alpha(dy) \\ = \ &\int_{M_\alpha} \{ \int_{N_y M^{r/\varepsilon}_\alpha} \phi(y + \varepsilon\, u) \ \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(y+\varepsilon\, u)\} \, p(y+\varepsilon\, u) \, J_\alpha(y,\varepsilon\, u) \, \lambda_{\alpha,y}(du) \ \} \, \lambda_\alpha(dy) \ , \end{aligned}$$

after rescaling the innermost integral. Notice that the factor ε^{-k} was chosen in order to compensate for this rescaling. The only indetermination left in the rescaled integral is with the exponential. We therefore focus on the expression

$$\frac{1}{\varepsilon^2}\,\ell^\varepsilon(y+\varepsilon\,u) = \tfrac{1}{2}\int_0^T |\frac{m_t(y+\varepsilon\,u)-m_t(y)}{\varepsilon}|^2\,dt - \int_0^T [\frac{m_t(y+\varepsilon\,u)-m_t(y)}{\varepsilon}]^*\,dW_t^\alpha \; . \tag{8}$$

Taylor expansion to the first order yields

$$\frac{1}{\varepsilon^2}\,\ell^\varepsilon(y+\varepsilon\,u) \longrightarrow Q_\alpha(y,u) = \tfrac{1}{2}\,u^*\,I_\perp(y)\,u - u^*\,\zeta_\alpha(y) \; ,$$

as $\varepsilon \downarrow 0$, where the Gaussian r.v. $\zeta_\alpha(y)$ has been defined in (4). Completing the square yields

$$Q_\alpha(y,u) = \tfrac{1}{2}\,(u-\chi_\alpha(y))^*\;I_\perp(y)\;(u-\chi_\alpha(y)) - \tfrac{1}{2}\,|\xi_\alpha(y)|^2 \; ,$$

where the Gaussian r.v.'s $\xi_\alpha(y)$ and $\chi_\alpha(y)$ have been defined in (5), hence

$$\exp\{-Q_\alpha(y,u)\}\,\lambda_{\alpha,y}(du) = (2\,\pi)^{k/2}\;\frac{\exp\{\frac{1}{2}\,|\xi_\alpha(y)|^2\}}{\sqrt{\det I_\perp(y)}}\;\Gamma_\alpha(y,du) \; , \tag{9}$$

where $\Gamma_\alpha(y,du)$ is a Gaussian probability distribution on the linear subspace N_yM_α, with *random* mean vector $\chi_\alpha(y)$ (a Gaussian r.v. with values in N_yM_α) and covariance matrix $[I_\perp(y)]^{-1}$. As a result, the following convergence results hold

$$\exp\{-\frac{1}{\varepsilon^2}\,\ell^\varepsilon(y+\varepsilon\,u)\}\,p(y+\varepsilon\,u)\,\lambda_{\alpha,y}(du)\,\lambda_\alpha(dy) \longrightarrow \Gamma_\alpha(y,du)\;\bar\mu_\alpha(dy) \; ,$$

and

$$\langle\bar\mu^\varepsilon,\phi\rangle \longrightarrow \langle\bar\mu_\alpha,\phi\rangle \; ,$$

as $\varepsilon \downarrow 0$. Further details are given in Appendix C. □

5 Rate of convergence

To study the rate of convergence, for $r_0 > 0$ small enough, fixed independently of ε, we define the projection π on M_α by

$$\pi(\theta) = \begin{cases} y \; , & \text{if } \theta = y+u \text{ with } (y,u)\in N^{r_0}M_\alpha, \\ \theta \; , & \text{otherwise,} \end{cases}$$

and we consider the small noise asymptotics of the joint conditional probability distribution ρ^ε of the r.v. $(\theta,\pi(\theta),\frac{1}{\varepsilon}\,[\theta-\pi(\theta)\,])$, defined by

$$\langle\rho^\varepsilon,f\rangle = \mathbf{E}^\varepsilon[f(\theta,\pi(\theta),\frac{\theta-\pi(\theta)}{\varepsilon})\mid\mathcal{Y}] \; .$$

Notice that the marginal w.r.t. the first variable is μ^ε, the marginal w.r.t. the second variable is the conditional probability distribution $\mu^\varepsilon\circ\pi^{-1}$ of the r.v. $\pi(\theta)$, and the marginal w.r.t. the last two variables is the joint conditional probability distribution ν^ε of the r.v. $(\pi(\theta),\frac{1}{\varepsilon}\,[\theta-\pi(\theta)\,])$, or equivalently the image of μ^ε under the mapping $\theta\mapsto(\pi(\theta),\frac{1}{\varepsilon}\,[\theta-\pi(\theta)\,])$, whose restriction to $M_\alpha^{r_0}$ takes values in the normal bundle space NM_α. Following the same approach as in Ellis and Rosen [5], we define the following *random* probability distribution

$$\langle\rho_\alpha,f\rangle = \int_{M_\alpha}\{\int_{N_yM_\alpha} f(y,y,u)\,\Gamma_\alpha(y,du)\;\}\;\mu_\alpha(dy) \; .$$

Notice that the marginal w.r.t. either the first or the second variable is μ_α, and the marginal w.r.t. the last two variables is the probability distribution ν_α on $N M_\alpha$ defined by

$$\langle \nu_\alpha, \psi \rangle = \int_{M_\alpha} \{ \int_{N_y M_\alpha} \psi(y,u)\, \Gamma_\alpha(y,du) \}\ \mu_\alpha(dy) \ ,$$

which is a mixture of *random* Gaussian probability distributions on the normal spaces.

The following norms are introduced : For any test function f defined on $\Theta \times \Theta \times \mathbb{R}^p$, let

$$\|f\|_W = \sup_{(\theta,y)\in\Theta, u\in\mathbb{R}^p} \frac{|f(\theta,y,u)|}{W(|u|)} \ , \qquad \|f\|_{\mathrm{L}(W)} = \sup_{(\theta,\theta',y)\in\Theta, u\in\mathbb{R}^p, \theta\neq\theta'} \frac{|f(\theta,y,u) - f(\theta',y,u)|}{|\theta-\theta'|\, W(|u|)} \ ,$$

and

$$\|f\|_{\mathrm{BL}(W)} = \max\{ \|f\|_W, \|f\|_{\mathrm{L}(W)} \} \ ,$$

where the enveloppe W is a nondecreasing real–valued function defined on $[0,\infty)$, and taking values larger than 1. A typical example is $W(r) = \max(1, r^q)$ for some $q \geq 0$. The corresponding dual norms are defined by

$$\|\rho - \rho'\|_W^* = \sup_{\|f\|_W = 1} |\, \langle \rho, f\rangle - \langle \rho', f\rangle \,| \qquad \text{and} \qquad \|\rho - \rho'\|_{\mathrm{BL}(W)}^* = \sup_{\|f\|_{\mathrm{BL}(W)} = 1} |\, \langle \rho, f\rangle - \langle \rho', f\rangle \,| \ .$$

Remark 5.1 If the test function f depends only upon the first variable, which is the situation of Theorem 4.1, i.e. if $f(\theta,y,u) = \phi(\theta)$, then

$$\|f\|_{\mathrm{BL}(W)} = \|\phi\|_{\mathrm{BL}} \qquad \text{hence} \qquad \|\mu^\varepsilon - \mu_\alpha\|_{\mathrm{BL}}^* \leq \|\rho^\varepsilon - \rho_\alpha\|_{\mathrm{BL}(W)}^* \ .$$

If the test function f depends only upon the second variable, i.e. if $f(\theta,y,u) = \phi(y)$, then

$$\|f\|_{\mathrm{BL}(W)} = \|\phi\| \qquad \text{hence} \qquad \|\mu^\varepsilon \circ \pi^{-1} - \mu_\alpha\|_{\mathrm{TV}} \leq \|\rho^\varepsilon - \rho_\alpha\|_{\mathrm{BL}(W)}^* \ .$$

Finally, if the test function f depends only upon the last two variables, i.e. if $f(\theta,y,u) = \psi(y,u)$, then

$$\|f\|_{\mathrm{BL}(W)} = \|\psi\|_W \qquad \text{hence} \qquad \|\nu^\varepsilon - \nu_\alpha\|_W^* \leq \|\rho^\varepsilon - \rho_\alpha\|_{\mathrm{BL}(W)}^* \ .$$

We introduce the following assumption on the enveloppe W, which is satisfied with $c_W = 1$, in the special case where $W(r) = \max(1, r^q)$ for some $q \geq 0$.

Assumption E : The enveloppe W is a nondecreasing real–valued function defined on $[0,\infty)$, taking values larger than 1, and such that :

(i) the following *integrability* condition holds

$$\frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} |v|^2\ W(|v|)\ \exp\{-\tfrac{1}{2}\,|v|^2\}\, dv < \infty \ ,$$

(ii) for any $\lambda > 0$ large enough

$$\lim_{r\uparrow\infty} r^k\, W(r)\ \exp\{-\lambda\, r\} \longrightarrow 0 \ ,$$

(iii) for any $r \geq 0$ and $r' \geq 0$

$$W(r\, r') \leq c_W\, W(r)\, W(r') \ .$$

Remark 5.2 Under part (i) of Assumption E

$$\frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} W(|v|)\ \exp\{-\tfrac{1}{2}\,|v|^2\}\, dv \leq W(1) + \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} |v|^2\ W(|v|)\ \exp\{-\tfrac{1}{2}\,|v|^2\}\, dv < \infty \ .$$

Theorem 5.3 *Under Assumptions A to E, and if the prior density p is positive on M_α, then*

$$\|\rho^\varepsilon - \rho_\alpha\|_{\mathrm{BL}(W)} \longrightarrow 0 \ ,$$

in $\mathbf{P}_\alpha^\varepsilon$–probability as $\varepsilon \downarrow 0$.

Remark 5.4 If $W \equiv 1$, then the following exponential rate is achieved : for any $\eta > 0$

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\kappa(\varepsilon)} \ \log \mathbf{P}_\alpha^\varepsilon [\, \|\rho^\varepsilon - \rho_\alpha\|_{\mathrm{BL}}^* > \eta] < 0 \ ,$$

where $\kappa(\varepsilon) \sim \varepsilon^{-2/5}\ \delta(\varepsilon)$, and $\delta(\varepsilon)$ goes to zero arbitrarily slowly as $\varepsilon \downarrow 0$.
If the enveloppe W is of the form $W(r) = \max(1, r^q)$ for some $q > 0$, and if in addition the prior density p is Lipschitz continuous on M_α, then the following exponential rate is achieved : for any $\eta > 0$

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\kappa(\varepsilon)} \ \log \mathbf{P}_\alpha^\varepsilon [\, \|\rho^\varepsilon - \rho_\alpha\|_{\mathrm{BL}(W)}^* > \eta] < 0 \ ,$$

where $\kappa(\varepsilon) \sim \varepsilon^{-2/(2q+5)}\ \delta(\varepsilon)$, and $\delta(\varepsilon)$ goes to zero arbitrarily slowly as $\varepsilon \downarrow 0$.

It follows from Remark 5.1 that Theorem 4.1 and Corollary 5.5 below can be obtained as special cases of Theorem 5.3.

Corollary 5.5 *Under the same assumptions as in Theorem 5.3*

$$\|\mu^\varepsilon \circ \pi^{-1} - \mu_\alpha\|_{\mathrm{TV}} \longrightarrow 0 \qquad \textit{and} \qquad \|\nu^\varepsilon - \nu_\alpha\|_W \longrightarrow 0 \ ,$$

in $\mathbf{P}_\alpha^\varepsilon$*–probability as* $\varepsilon \downarrow 0$.

Remark 5.6 If the test function f is of the special form $f(\theta, y, u) = u$, then

$$\int_{N_y M_\alpha} u\, \Gamma_\alpha(y, du) = \chi_\alpha(y) \ ,$$

and Theorem 5.3 implies that

$$|\, \mathbf{E}^\varepsilon [\frac{\theta - \pi(\theta)}{\varepsilon} \mid \mathcal{Y}] - \int_{M_\alpha} \chi_\alpha(y)\, \mu_\alpha(dy) \mid \longrightarrow 0 \ ,$$

in $\mathbf{P}_\alpha^\varepsilon$–probability as $\varepsilon \downarrow 0$.

Another approach to look at rate of convergence, would be to consider the asymptotic behavior of the normalized difference $\dfrac{1}{\varepsilon}\, [\mu^\varepsilon - \mu_\alpha]$ as $\varepsilon \downarrow 0$. This problem will be considered in a subsequent paper.

6 Application to TMA

Target motion analysis (TMA), or tracking with bearings only measurements, is a typical application where nonobservability occurs, see Lévine and Marino [17]. In this problem, a ship (the platform) is trying to estimate the position and velocity of another ship (the target), using angle only measurements provided by a passive sonar. We denote by $r = (r^x, r^y)$ and $v = (v^x, v^y)$ the relative position and velocity of the target w.r.t. the platform. The available measurements are of the form

$$dY_t = h(r_t, v_t)\, dt + \varepsilon\, dW_t$$

where the observation function is

$$h(r, v) = \arctan \frac{r^x}{r^y} \ .$$

In general, both ships move at constant velocity along straight lines, so that the motion of the target could be described by the equation

$$r_t^x = r_0^x + v_0^x\, t \ , \qquad r_t^y = r_0^y + v_0^y\, t \ ,$$

$$v_t^x = v_0^x \ , \qquad v_t^y = v_0^y \ .$$

Denoting by $\theta = (r_0^x, r_0^y, v_0^x, v_0^y)$ the initial condition of the above set of equations, the problem reduces to estimating θ based on the observations

$$dY_t = m_t(\theta)\, dt + \varepsilon\, dW_t^\theta \ ,$$

where

$$m_t(\theta) = \arctan \frac{\theta_1 + t\,\theta_3}{\theta_2 + t\,\theta_4} \ .$$

Denote by $M_\alpha = \{\theta \in \mathbb{R}^4 \ : \ m_t(\theta) = m_t(\alpha)\, ,\, 0 \leq t \leq T\}$ the set of points that cannot be distinguished from α. Clearly, $\theta \in M_\alpha$ if and only if for any $0 \leq t \leq T$

$$\frac{\theta_1 + t\,\theta_3}{\theta_2 + t\,\theta_4} = \frac{\alpha_1 + t\,\alpha_3}{\alpha_2 + t\,\alpha_4} \ ,$$

i.e. if and only if

$$\begin{pmatrix} \alpha_2 & -\alpha_1 & 0 & 0 \\ \alpha_4 & -\alpha_3 & \alpha_2 & -\alpha_1 \\ 0 & 0 & \alpha_4 & -\alpha_3 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} = 0 \ .$$

Notice that the rows of this matrix are linearly independent vectors of $\mathbb{R}^4$, provided that $\alpha_4\,\alpha_1 - \alpha_2\,\alpha_3 \neq 0$. If this is the case, then M_α is the one–dimensional kernel of this matrix. We can easily check that α is in the kernel, hence

$$M_\alpha = \{y \in \mathbb{R}^4 \ : \ y = \rho\,\alpha \quad \text{for some } \rho > 0\} \ .$$

Indeed there exists a one–dimensional manifold, in the four–dimensional parameter space (initial relative position and velocity of the target w.r.t. the platform), of points that cannot be distinguished from the *true* value. One would need additional measurements, e.g. range measurements provided by an active sonar, to overcome this problem. The case where $\alpha_4\,\alpha_1 - \alpha_2\,\alpha_3 = 0$ corresponds to some values of α for which the platform and the target are moving along the same line. We do not consider this type of problem here and therefore, we discard these values of α from the set of parameter Θ.

Simple calculations yield

$$[\dot{m}_t(\theta)]^* = \frac{1}{[\theta_1 + t\,\theta_3]^2 + [\theta_2 + t\,\theta_4]^2} \begin{pmatrix} \theta_2 + t\,\theta_4 \\ -[\theta_1 + t\,\theta_3] \\ t\,[\theta_2 + t\,\theta_4] \\ -t\,[\theta_1 + t\,\theta_3] \end{pmatrix} \ ,$$

for any $\theta \in \Theta$, hence for any $\rho > 0$

$$\dot{m}_t(\rho\,\alpha) = \frac{1}{\rho}\,\dot{m}_t(\alpha) \ .$$

It follows that for any $\rho > 0$

$$I(\rho\,\alpha) = \int_0^T [\dot{m}_t(\rho\,\alpha)]^*\,\dot{m}_t(\rho\,\alpha)\,dt = \frac{1}{\rho^2}\,I(\alpha) \ ,$$

and similarly for the 3×3 matrix

$$I_\perp(\rho\,\alpha) = \frac{1}{\rho^2}\,I_\perp(\alpha) \ ,$$

hence

$$\sqrt{\det I_\perp(\rho\,\alpha)} = \frac{1}{\rho^3}\,\sqrt{\det I_\perp(\alpha)} \ .$$

It follows also that for any $\rho > 0$

$$\xi_\alpha(\rho\,\alpha) = [I_\perp(\rho\,\alpha)]^{-1/2} \int_0^T [\dot m_t(\rho\,\alpha)]^* \, dW_t^\alpha = \xi_\alpha(\alpha) \; ,$$

does not depend on ρ, i.e. is constant over M_α, and

$$\chi_\alpha(\rho\,\alpha) = [I_\perp(\rho\,\alpha)]^{-1} \int_0^T [\dot m_t(\rho\,\alpha)]^* \, dW_t^\alpha = \rho\,\chi_\alpha(\alpha) \; .$$

Even though the parameter space is not compact in this example, the results obtained in the previous section are still valid, up to straightforward modifications of the proofs. The limiting probability distribution μ_α on M_α has the form

$$\langle \mu_\alpha, \phi \rangle = c_\alpha \int_0^\infty \phi(\rho\,\alpha)\, \rho^3 \, p(\rho\,\alpha)\, d\rho = \int_0^\infty \phi(\rho\,\alpha)\, \gamma_\alpha(d\rho) \; ,$$

where c_α is a normalizing constant, i.e. the probability distribution μ_α on M_α is the image of the probability distribution γ_α on $[0,\infty)$ under the (parametrisation) mapping $\rho \mapsto y = \rho\,\alpha$. Notice that in this example μ_α is *nonrandom*, and does not depend on the observation duration T.

Since M_α is a one–dimensional *linear* submanifold, the normal spaces $N_y M_\alpha$ for $y \in M_\alpha$ are three–dimensional linear submanifolds, all parallel to the single vector space

$$M_\alpha^\perp = \{ v \in \mathbb{R}^4 \; : \; v^* \, \alpha = 0 \} \; ,$$

and the probability distribution governing the rate of convergence has the form

$$\begin{aligned}
\langle \nu_\alpha, \psi \rangle &= \int_0^\infty \frac{\sqrt{\det I_\perp(\rho\,\alpha)}}{(2\,\pi)^{3/2}} \int_{M_\alpha^\perp} \psi(\rho\,\alpha, u)\, \exp\{-\tfrac{1}{2}\, [u - \chi_\alpha(\rho\,\alpha)]^* \, I_\perp(\rho\,\alpha)\, [u - \chi_\alpha(\rho\,\alpha)] \,\}\, du \; \gamma_\alpha(d\rho) \\
&= \int_0^\infty \frac{1}{\rho^3}\, \frac{\sqrt{\det I_\perp(\alpha)}}{(2\,\pi)^{3/2}} \int_{M_\alpha^\perp} \psi(\rho\,\alpha, u)\, \exp\{-\tfrac{1}{2}\, [u - \rho\,\chi_\alpha(\alpha)]^* \, \frac{1}{\rho^2}\, I_\perp(\alpha)\, [u - \rho\,\chi_\alpha(\alpha)] \,\}\, du \; \gamma_\alpha(d\rho) \\
&= \int_0^\infty \frac{\sqrt{\det I_\perp(\alpha)}}{(2\,\pi)^{3/2}} \int_{M_\alpha^\perp} \psi(\rho\,\alpha, \rho\, v)\, \exp\{-\tfrac{1}{2}\, [v - \chi_\alpha(\alpha)]^* \, I_\perp(\alpha)\, [v - \chi_\alpha(\alpha)]\}\, dv \; \gamma_\alpha(d\rho) \; .
\end{aligned}$$

In the special case where $\psi(y,u) = u$

$$|\, \mathbf{E}^\varepsilon[\, \frac{\theta - \pi(\theta)}{\varepsilon} \mid \mathcal{Y}] - \chi_\alpha(\alpha)\, \frac{\displaystyle\int_0^\infty \rho^4 \, p(\rho\,\alpha)\, d\rho}{\displaystyle\int_0^\infty \rho^3 \, p(\rho\,\alpha)\, d\rho} \,| \longrightarrow 0 \; ,$$

in $\mathbf{P}_\alpha^\varepsilon$–probability as $\varepsilon \downarrow 0$.

Acknowledgment

The authors gratefully acknowledge Matthew James for bringing the papers [5, 6] to their attention, and Fabien Campillo for stimulating their interest in this problem.

References

[1] M. Berger and B. Gostiaux. *Differential Geometry : Manifolds, Curves and Surfaces*, volume 115 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, 1988.

[2] V. S. Borkar and A. Bagchi. Parameter estimation in continuous–time stochastic processes. *Stochastics*, 8(3):193–212, 1982.

[3] R. M. Dudley. Convergence of Baire measures. *Studia Mathematica*, XXVII:251–268, 1966.

[4] R. M. Dudley. Distance of probability measures and random variables. *The Annals of Mathematical Statistics*, 39(5):1563–1572, Oct. 1968.

[5] R. S. Ellis and J. S. Rosen. Asymptotic analysis of Gaussian integrals, II : Manifold of minimum points. *Communications in Mathematical Physics*, 82(2):153–181, 1981.

[6] R. S. Ellis and J. S. Rosen. Asymptotic analysis of Gaussian integrals. I : Isolated minimum points. *Transactions of the American Mathematical Society*, 273(2):447–481, Oct. 1982.

[7] Hwang Chii–Ruey. Laplace's method revisited : weak convergence of probability measures. *The Annals of Probability*, 8(6):1177–1182, Dec. 1980.

[8] I. A. Ibragimov and R. Z. Khasminskii. *Statistical Estimation. Asymptotic Theory*, volume 16 of *Applications of Mathematics*. Springer–Verlag, New York, 1981.

[9] M. R. James. Finite time observer design by probabilistic–variational methods. *SIAM Journal on Control and Optimization*, 29(4):954–967, July 1991.

[10] M. R. James and F. Le Gland. Consistent parameter estimation for partially observed diffusions with small noise. *Applied Mathematics & Optimization*, 32(1):47–72, July/Aug. 1995.

[11] M. Joannides and F. Le Gland. Small noise asymptotics of the Bayesian estimator in nonidentifiable nonlinear regressions. In *Proceedings of the 2nd Portuguese Conference on Automatic Control, Porto 1996*, pages 257–261. APCA, Sept. 1996.

[12] M. Joannides and F. Le Gland. Small noise asymptotics of nonlinear filters with nonobservable limiting deterministic system. In *Proceedings of the 36th Conference on Decision and Control, San Diego 1997*, pages 1663–1668. IEEE–CSS, Dec. 1997.

[13] Yu. A. Kutoyants. *Identification of Dynamical Systems with Small Noise*, volume 300 of *Mathematics and its Applications*. Kluwer Academic Publisher, Dordrecht, 1994.

[14] Yu. A. Kutoyants. Nonconsistent estimation by diffusion type observations. *Statistics & Probability Letters*, 20(1):1–7, May 1994.

[15] Yu. A. Kutoyants and L. Yu. Vostrikova. On non–consistency of estimators. *Stochastics and Stochastics Reports*, 53(1):53–80, 1995.

[16] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer–Verlag, Berlin, 1991.

[17] J. Lévine and R. Marino. Constant–speed target tracking via bearings–only measurements. *IEEE Transactions on Aerospace and Electronic Systems*, AES–28(1):174–182, Jan. 1992.

[18] M. B. Marcus and L. A. Shepp. Sample behavior of Gaussian processes. In L. Le Cam, J. Neyman, and E. Scott, editors, *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 423–441. University of California Press, 1972.

[19] H. Weyl. On the volume of tubes. *American Journal of Mathematics*, 61:461–472, 1939.

A Proof of Proposition 3.2

Notice that for any $\theta \in \Theta$

$$-\ell^\varepsilon(\theta) = -K_\alpha(\theta) + \varepsilon \int_0^T [m_t(\theta) - m_t(\alpha)]^* \, dW_t^\alpha \ ,$$

hence

$$-K_\alpha(\theta) - \varepsilon \, S_0 \leq -\ell^\varepsilon(\theta) \leq -K_\alpha(\theta) + \varepsilon \, S_0 \ ,$$

where the r.v.

$$S_0 = \sup_{\theta \in \Theta} | \int_0^T [m_t(\theta) - m_t(\alpha)]^* \, dW_t^\alpha \ | \ ,$$

is $\mathbf{P}_\alpha^\varepsilon$–a.s. finite for any $\varepsilon > 0$, under Assumption A. Following the same approach as in Hwang [7, Section 2]

$$\begin{aligned}
\mu^\varepsilon(K_\alpha(\theta) \geq c) &= \frac{\displaystyle\int_\Theta \mathbf{1}_{(K_\alpha(\theta) \geq c)} \ \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(\theta)\} \, \mu(d\theta)}{\displaystyle\int_\Theta \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(\theta)\} \, \mu(d\theta)} \\
&\leq \frac{\displaystyle\int_\Theta \mathbf{1}_{(K_\alpha(\theta) \geq c)} \ \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(\theta)\} \, \mu(d\theta)}{\displaystyle\int_\Theta \mathbf{1}_{(K_\alpha(\theta) \leq \frac{1}{2} c)} \ \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(\theta)\} \, \mu(d\theta)} \leq \frac{\exp\{\dfrac{2 \, S_0}{\varepsilon}\} \ \exp\{-\dfrac{c}{2 \, \varepsilon^2}\}}{\displaystyle\int_\Theta \mathbf{1}_{(K_\alpha(\theta) \leq \frac{1}{2} c)} \, \mu(d\theta)} \ .
\end{aligned}$$

Introducing the *bad* set

$$\Omega_0 = \{ \sup_{\theta \in \Theta} | \int_0^T [m_t(\theta) - m_t(\alpha)]^* \, dW_t^\alpha \, | > \frac{c}{8 \, \varepsilon} \ \} \ ,$$

the following estimate holds : On the *good* set $\Omega \setminus \Omega_0$

$$\mu^\varepsilon(K_\alpha(\theta) \geq c) \leq \frac{\exp\{-\dfrac{c}{4 \, \varepsilon^2}\}}{\mu(K_\alpha(\theta) \leq \frac{1}{2} c)} \ . \tag{10}$$

For any $\eta > 0$

$$\mathbf{P}_\alpha^\varepsilon[\mu^\varepsilon(K_\alpha(\theta) \geq c) > \eta] \leq \mathbf{P}_\alpha^\varepsilon[\mu^\varepsilon(K_\alpha(\theta) \geq c) > \eta \, , \, \Omega \setminus \Omega_0] + \mathbf{P}_\alpha^\varepsilon[\Omega_0] \ ,$$

and from estimate (10) it follows that

$$\mathbf{P}_\alpha^\varepsilon[\mu^\varepsilon(K_\alpha(\theta) \geq c) > \eta \, , \, \Omega \setminus \Omega_0] = 0 \ ,$$

for $\varepsilon > 0$ small enough. On the other hand, $\mathbf{P}_\alpha^\varepsilon[\Omega_0]$ goes to zero as $\varepsilon \downarrow 0$.

B Proof of Lemma 3.4

It follows from Remark 3.1 that $\lambda_{\max} < \infty$ under Assumption A. To check that $\lambda_{\min} > 0$, assume that there exists a sequence $(y_n, u_n) \in N^1 M_\alpha$ such that

$$u_n^* \, I_\perp(y_n) \, u_n \leq 1/n \ .$$

Since $N^1 M_\alpha$ is compact, there exists a subsequence which converges to a limit $(y_*, u_*) \in N^1 M_\alpha$. Taking the limit along this subsequence, yields

$$u_*^* \, I_\perp(y_*) \, u_* = 0 \ ,$$

by continuity, which contradicts $\lambda_{\min}(y_*) > 0$.

▶ **Lower bound :** For any $r > 0$ define

$$g_\alpha(r) = \inf_{\theta \in \Theta \setminus M_\alpha^r} K_\alpha(\theta) ,$$

which is positive since $\Theta \setminus M_\alpha^r$ is compact. Notice that for any $r' > r$

$$g_\alpha(r) = \min \{ \, g_\alpha(r') \, , \inf_{\theta \in M_\alpha^{r'} \setminus M_\alpha^r} K_\alpha(\theta) \, \} .$$

By the tubular neighborhood theorem, see Berger and Gostiaux [1, Theorem 2.7.12], if r' is small enough, then any $\theta \in M_\alpha^{r'} \setminus M_\alpha^r$ can be written as $\theta = y + u$ with $(y,u) \in N\,M_\alpha$ and $r \leq |u| < r'$. Taylor expansion yields

$$K_\alpha(\theta) = \int_0^1 (1-s)\, \ddot{K}_\alpha(y + s\,u)\,(u,u)\,ds \geq \tfrac{1}{2}\, r^2 \inf_{(y,v) \in N^1 M_\alpha} \inf_{0 \leq s \leq r'} \ddot{K}_\alpha(y + s\,v)\,(v,v) .$$

There exists $r_{\min} > 0$ such that for any $(y,v) \in N^1 M_\alpha$ and any $0 \leq s \leq r_{\min}$

$$\ddot{K}_\alpha(y + s\,v)\,(v,v) \geq \tfrac{1}{2}\, \lambda_{\min} .$$

Indeed, assume that there exist sequences $(y_n, v_n) \in N^1 M_\alpha$ and $0 \leq s_n \leq 1/n$ such that

$$\ddot{K}_\alpha(y_n + s_n\, v_n)\,(v_n, v_n) < \tfrac{1}{2}\, \lambda_{\min} .$$

Since $N^1 M_\alpha$ is compact, there exists a subsequence which converges to a limit $(y_*, v_*) \in N^1 M_\alpha$. Taking the limit along this subsequence, yields

$$\ddot{K}_\alpha(y_*)\,(v_*, v_*) = v_*^*\, I_\perp(y_*)\, v_* \leq \tfrac{1}{2}\, \lambda_{\min} ,$$

by continuity, which contradicts the definition of $\lambda_{\min}$. Therefore

$$\inf_{\theta \in M_\alpha^{r_{\min}} \setminus M_\alpha^r} K_\alpha(\theta) \geq \tfrac{1}{4}\, \lambda_{\min}\, r^2 ,$$

for any $0 < r < r_{\min}$, hence

$$g_\alpha(r) \geq \min \{ \, g_\alpha(r_{\min}) \, , \tfrac{1}{4}\, \lambda_{\min}\, r^2 \, \} ,$$

and for $r > 0$ small enough

$$g_\alpha(r) \geq \tfrac{1}{4}\, \lambda_{\min}\, r^2 .$$

▶ **Upper bound :** For any $\theta \in \Theta$, and any $y \in M_\alpha$, Taylor expansion yields

$$K_\alpha(\theta) = \int_0^1 (1-s)\, \ddot{K}_\alpha(y + s\,(\theta - y))\,(\theta - y, \theta - y)\,ds .$$

If $\theta \in M_\alpha^r$, then since M_α is compact there exists $y(\theta) \in M_\alpha$ such that $|\theta - y(\theta)| \leq r$, hence

$$K_\alpha(\theta) \leq \tfrac{1}{2}\, r^2 \sup_{|u|=1} \sup_{0 \leq s \leq r} \ddot{K}_\alpha(y(\theta) + s\,u)\,(u,u) \leq \tfrac{1}{2}\, r^2 \sup_{y \in M_\alpha} \sup_{|u|=1} \sup_{0 \leq s \leq r} \ddot{K}_\alpha(y + s\,u)\,(u,u) .$$

There exists $r_{\min} > 0$ such that for any $(y,u) \in M_\alpha \times \mathbf{S}^{p-1}$ and any $0 \leq s \leq r_{\min}$

$$\ddot{K}_\alpha(y + s\,u)\,(u,u) \leq 2\, \lambda_{\max} .$$

Indeed, assume that there exist sequences $(y_n, u_n) \in M_\alpha \times \mathbf{S}^{p-1}$ and $0 \leq s_n \leq 1/n$ such that

$$\ddot{K}_\alpha(y_n + s_n\, u_n)\,(u_n, u_n) > 2\, \lambda_{\max} .$$

Since $M_\alpha \times \mathbf{S}^{p-1}$ is compact, there exists a subsequence which converges to a limit $(y_*, u_*) \in M_\alpha \times \mathbf{S}^{p-1}$. Taking the limit along this subsequence, yields

$$\ddot{K}_\alpha(y_*)\,(u_*, u_*) = u_*^*\, I(y_*)\, u_* \geq 2\, \lambda_{\max} ,$$

by continuity, which contradicts the definition of $\lambda_{\max}$. Therefore

$$\sup_{\theta \in M_\alpha^r} K_\alpha(\theta) \leq \lambda_{\max}\, r^2 ,$$

for any $0 < r < r_{\min}$.

C Proof of Theorem 5.3

Notice that $|\theta - \pi(\theta)| \leq r_0$ for any $\theta \in \Theta$, hence

$$|f(\theta, \pi(\theta), \frac{\theta - \pi(\theta)}{\varepsilon})| \leq W(\frac{r_0}{\varepsilon})\; \|f\|_W \;,$$

and for any $\theta = y + \varepsilon\, u$ with $(y,u) \in N^{r_0/\varepsilon} M_\alpha$

$$f(\theta, \pi(\theta), \frac{\theta - \pi(\theta)}{\varepsilon}) = f(y + \varepsilon\, u, y, u) \;.$$

As in the (sketch of the) proof of Theorem 4.1, we introduce the unnormalized probability distributions $\bar{\rho}^\varepsilon$ and $\bar{\rho}_\alpha$ defined by

$$\langle \bar{\rho}^\varepsilon, f \rangle = \varepsilon^{-k} \int_\Theta f(\theta, \pi(\theta), \frac{\theta - \pi(\theta)}{\varepsilon})\; \exp\{-\frac{1}{\varepsilon^2}\, \ell^\varepsilon(\theta)\}\, p(\theta)\, d\theta \;,$$

and

$$\langle \bar{\rho}_\alpha, f \rangle = (2\,\pi)^{k/2} \int_{M_\alpha} \{ \int_{N_y M_\alpha} f(y,y,u)\, \Gamma_\alpha(y, du) \}\; \frac{\exp\{\frac{1}{2}\, |\xi_\alpha(y)|^2\}}{\sqrt{\det I_\perp(y)}}\; p(y)\, \lambda_\alpha(dy) \;,$$

respectively, with normalizing constants $\langle \bar{\rho}^\varepsilon, \mathbf{1} \rangle = \langle \bar{\mu}^\varepsilon, \mathbf{1} \rangle$ and $\langle \bar{\rho}_\alpha, \mathbf{1} \rangle = \langle \bar{\mu}_\alpha, \mathbf{1} \rangle$.

▶ **Localization :** The first approximation is to replace the integral over the whole space Θ with the integral over the tubular neighborhood M_α^r, i.e. to define

$$\langle \bar{\rho}_{\rm I}, f \rangle = \varepsilon^{-k} \int_{M_\alpha^r} f(\theta, \pi(\theta), \frac{\theta - \pi(\theta)}{\varepsilon})\; \exp\{-\frac{1}{\varepsilon^2}\, \ell^\varepsilon(\theta)\}\, p(\theta)\, d\theta \;,$$

and the corresponding error $\delta_{\rm I} = \bar{\rho}^\varepsilon - \bar{\rho}_{\rm I}$. It holds

$$|\langle \delta_{\rm I}, f \rangle| \leq \varepsilon^{-k}\, W(\frac{r_0}{\varepsilon})\; \|f\|_W \int_{\Theta \setminus M_\alpha^r} \exp\{-\frac{1}{\varepsilon^2}\, \ell^\varepsilon(\theta)\}\, p(\theta)\, d\theta \;.$$

For any $r > 0$ define

$$g_\alpha(r) = \inf_{\theta \in \Theta \setminus M_\alpha^r} K_\alpha(\theta) \;,$$

which is positive since $\Theta \setminus M_\alpha^r$ is compact. Notice that for any $\theta \in \Theta \setminus M_\alpha^r$

$$-\ell^\varepsilon(\theta) = -K_\alpha(\theta) + \varepsilon \int_0^T [m_t(\theta) - m_t(\alpha)]^*\, dW_t^\alpha \leq -g_\alpha(r) + \varepsilon\, S_0 \;,$$

where the r.v.

$$S_0 = \sup_{\theta \in \Theta} |\int_0^T [m_t(\theta) - m_t(\alpha)]^*\, dW_t^\alpha\, | \;,$$

is $\mathbf{P}_\alpha^\varepsilon$–a.s. finite for any $\varepsilon > 0$, under Assumption A. Introducing the *bad* set

$$\Omega_{\rm I} = \{ \sup_{\theta \in \Theta} |\int_0^T [m_t(\theta) - m_t(\alpha)]^*\, dW_t^\alpha\, | > \frac{g_\alpha(r)}{2\,\varepsilon} \} \;,$$

the following estimate holds : On the *good* set $\Omega \setminus \Omega_{\rm I}$

$$|\langle \delta_{\rm I}, f \rangle| \leq \varepsilon^{-k}\, W(\frac{r_0}{\varepsilon})\; \exp\{-\frac{g_\alpha(r)}{2\,\varepsilon^2}\}\; \|f\|_W \;,$$

and it follows from Lemma 3.4 that, for $r > 0$ small enough

$$|\langle \delta_{\rm I}, f \rangle| \leq \varepsilon^{-k}\, W(\frac{r_0}{\varepsilon})\; \exp\{-\frac{r^2}{8\,\varepsilon^2}\, \lambda_{\min}\}\; \|f\|_W \;. \tag{11}$$

Notice that the normalizing constant $\langle \bar{\rho}_\alpha, \mathbf{1} \rangle = \langle \bar{\mu}_\alpha, \mathbf{1} \rangle$ is bounded from below by

$$\Lambda_\alpha = (2\,\pi)^{k/2} \int_{M_\alpha} \frac{p(y)}{\sqrt{\det I_\perp(y)}}\, \lambda_\alpha(dy) \ ,$$

hence estimate (11) gives, for $r > 0$ small enough

$$\frac{|\langle \delta_{\mathrm{I}}, f \rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1} \rangle} \leq \frac{1}{\Lambda_\alpha}\, \varepsilon^{-k}\, W(\frac{r_0}{\varepsilon})\ \exp\{-\frac{r^2}{8\,\varepsilon^2}\, \lambda_{\min}\}\ \|f\|_W \ .$$

▶ **Taylor expansion :** Expansion to the first order yields

$$\begin{aligned} \frac{m_t(y+\varepsilon\,u) - m_t(y)}{\varepsilon} &= \dot{m}_t(y)\,u + \varepsilon \int_0^1 (1-s)\, \ddot{m}_t(y+\varepsilon\,s\,u)(u,u)\,ds \\ &= \dot{m}_t(y)\,u + \varepsilon\, R_t(y, \varepsilon\,u)\,(u,u) \end{aligned}$$

where for any $(y,z) \in N^r M_\alpha$ the bilinear mapping $R_t(y,z)$ is defined on $N_y M_\alpha$ by

$$R_t(y,z)\,(u,u) = \int_0^1 (1-s)\, \ddot{m}_t(y+s\,z)\,(u,u)\,ds \ .$$

The mapping $u \longmapsto \dfrac{1}{\varepsilon^2}\, \ell^\varepsilon(y+\varepsilon\,u)$ defined on the normal space $N_y M_\alpha$ appears then as a small perturbation of the (nondegenerate) quadratic form $u \longmapsto Q_\alpha(y,u)$ defined on $N_y M_\alpha$ by

$$Q_\alpha(y,u) = \tfrac{1}{2}\, u^*\, I_\perp(y)\, u - u^*\, \zeta_\alpha(y) \ ,$$

where $I_\perp(y)$ and $\zeta_\alpha(y)$ have been defined in Section 3. Indeed, we can rewrite (8) as

$$\begin{aligned} \frac{1}{\varepsilon^2}\, \ell^\varepsilon(y+\varepsilon\,u) &= \tfrac{1}{2} \int_0^T [\, |\dot{m}_t(y)\,u|^2 + 2\,\varepsilon\,[\dot{m}_t(y)\,u]^*\, R_t(y,\varepsilon\,u)\,(u,u) + \varepsilon^2\, |R_t(y,\varepsilon\,u)\,(u,u)|^2\,]\,dt \\ &\qquad - \int_0^T [\, \dot{m}_t(y)\,u + \varepsilon\, R_t(y,\varepsilon\,u)\,(u,u)\,]^*\, dW_t^\alpha \\ &= Q_\alpha(y,u) + F_\alpha^\varepsilon(u,y,\varepsilon\,u) - \varepsilon\, H_\alpha(u,y,\varepsilon\,u) \ , \end{aligned}$$

where for any $(y,z) \in N^r M_\alpha$

$$F_\alpha^\varepsilon(u,y,z) = \varepsilon \int_0^T [\dot{m}_t(y)\,u]^*\, R_t(y,z)(u,u)\,dt + \tfrac{1}{2}\,\varepsilon^2 \int_0^T |R_t(y,z)(u,u)|^2\,dt \ ,$$

and

$$H_\alpha(u,y,z) = \int_0^T [R_t(y,z)\,(u,u)]^*\, dW_t^\alpha \ .$$

Recall that

$$\begin{aligned} \langle \bar{\rho}_{\mathrm{I}}, f \rangle &= \int_{M_\alpha} \int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\,|u| \leq r)}\, f(y+\varepsilon\,u, y, u)\, \exp\{-\frac{1}{\varepsilon^2}\, \ell^\varepsilon(y+\varepsilon\,u)\} \\ &\qquad\qquad p(y+\varepsilon\,u)\, J_\alpha(y,\varepsilon\,u)\, \lambda_{\alpha,y}(du)\, \lambda_\alpha(dy) \ . \end{aligned}$$

The second approximation is to replace $\dfrac{1}{\varepsilon^2}\, \ell^\varepsilon(y+\varepsilon\,u)$ by the quadratic form $Q_\alpha(y,u)$, i.e. to define

$$\begin{aligned} \langle \bar{\rho}_{\mathrm{II}}, f \rangle &= \int_{M_\alpha} \int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\,|u| \leq r)}\, f(y+\varepsilon\,u, y, u)\, \exp\{-Q_\alpha(y,u)\} \\ &\qquad\qquad p(y+\varepsilon\,u)\, J_\alpha(y,\varepsilon\,u)\, \lambda_{\alpha,y}(du)\, \lambda_\alpha(dy) \end{aligned} \tag{12}$$

and the corresponding error $\delta_{\rm II} = \bar\rho_{\rm I} - \bar\rho_{\rm II}$. It holds

$$\begin{aligned} |\langle \delta_{\rm II}, f\rangle| \quad \leq \quad & s_1(r)\, \|f\|_W \int_{M_\alpha} \int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\, |u| \leq r)}\, | \exp\{-F^\varepsilon_\alpha(u,y,\varepsilon\, u) + \varepsilon\, H_\alpha(u,y,\varepsilon\, u)\} - 1| \\ & \qquad W(|u|)\ \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\ p(y)\, \lambda_\alpha(dy)\ , \end{aligned} \tag{13}$$

where

$$s_1(r) = \sup_{(y,z) \in N^r M_\alpha} \frac{p(y+z)}{p(y)}\, J_\alpha(y,z)$$

is finite for small $r > 0$, since p and J_α are continuous and $N^r M_\alpha$ is compact.

- We first estimate the deterministic term in the remainder of the Taylor expansion. It holds

$$\int_0^T |R_t(y,z)(u,u)|^2\, dt \leq \tfrac{1}{4} \sup_{0 \leq s \leq 1} \int_0^T |\ddot m_t(y + s\, z)(u,u)|^2\, dt\ ,$$

hence

$$\begin{aligned} \sup_{(y,z) \in N^r M_\alpha} \int_0^T |R_t(y,z)(u,u)|^2\, dt \quad & \leq \quad \tfrac{1}{4} \sup_{(y,z) \in N^r M_\alpha} \sup_{0 \leq s \leq 1} \int_0^T |\ddot m_t(y + s\, z)(u,u)|^2\, dt \\ & \leq \quad \tfrac{1}{4} \sup_{\theta \in \Theta} \int_0^T |\ddot m_t(\theta)(u,u)|^2\, dt \leq \tfrac{1}{4}\, \sigma_2^2\, |u|^4\ , \end{aligned} \tag{14}$$

from Assumption B. On the other hand, we have

$$| \int_0^T [\dot m_t(y)\, u]^*\, R_t(y,z)(u,u)\, dt\, | \leq \{ \int_0^T |\dot m_t(y)\, u|^2\, dt\}^{1/2}\ \{ \int_0^T |R_t(y,z)(u,u)|^2\, dt\}^{1/2}\ ,$$

by the Cauchy–Schwartz inequality, hence using (14) and Assumption A, we get

$$| \int_0^T [\dot m_t(y)\, u]^*\, R_t(y,z)(u,u)\, dt\, | \leq \tfrac{1}{2}\, \sigma_1\, \sigma_2\, |u|^3\ .$$

Combining the above estimates yields

$$|F^\varepsilon_\alpha(u,y,\varepsilon\, u)| \leq \tfrac{1}{2}\, \sigma_1\, \sigma_2\, \varepsilon\, |u|^3 + \tfrac{1}{8}\, \sigma_2^2\, \varepsilon^2\, |u|^4 \leq \tfrac{1}{2}\, \sigma_1\, \sigma_2\, r\, |u|^2 + \tfrac{1}{8}\, \sigma_2^2\, r^2\, |u|^2 = \tfrac{1}{4}\, r\, s_2(r)\, |u|^2\ ,$$

for any $u \in N^{r/\varepsilon}_y M_\alpha$, where

$$\tfrac{1}{4}\, s_2(r) = \tfrac{1}{2}\, \sigma_1\, \sigma_2 + \tfrac{1}{8}\, \sigma_2^2\, r\ ,$$

is finite for small $r > 0$.

- We estimate now the random term in the remainder of the Taylor expansion. By definition

$$H_\alpha(u,y,z) = \int_0^1 (1-s) \int_0^T [\ddot m_t(y + s\, z)\, (u,u)]^*\, dW^\alpha_t\ ds\ ,$$

hence the following estimate holds

$$|H_\alpha(u,y,z)| \leq \tfrac{1}{2} \sup_{0 \leq s \leq 1} | \int_0^T [\ddot m_t(y + s\, z)\, (u,u)]^*\, dW^\alpha_t\, |\ ,$$

and

$$\begin{aligned} \sup_{(y,z) \in N^r M_\alpha} |H_\alpha(u,y,z)| \quad & \leq \quad \tfrac{1}{2} \sup_{(y,z) \in N^r M_\alpha} \sup_{0 \leq s \leq 1} | \int_0^T [\ddot m_t(y + s\, z)\, (u,u)]^*\, dW^\alpha_t\, | \\ & \leq \quad \tfrac{1}{2} \sup_{\theta \in M^r_\alpha} | \int_0^T [\ddot m_t(\theta)\, (u,u)]^*\, dW^\alpha_t\, | \\ & \leq \quad \tfrac{1}{2}\, |u|^2 \sup_{\theta \in M^r_\alpha} \sup_{|u|=1} | \int_0^T [\ddot m_t(\theta)\, (u,u)]^*\, dW^\alpha_t\, | \leq \tfrac{1}{2}\, |u|^2\, S_2\ , \end{aligned}$$

where the r.v.

$$S_2 = \sup_{\theta\in\Theta} \sup_{|u|=1} | \int_0^T [\ddot{m}_t(\theta)\,(u,u)]^*\, dW_t^\alpha \,|$$

is $\mathbf{P}_\alpha^\varepsilon$–a.s. finite for any $\varepsilon > 0$, under Assumption B. Then introducing the *bad* set

$$\Omega'_{\rm II} = \{ \sup_{\theta\in\Theta} \sup_{|u|=1} | \int_0^T [\ddot{m}_t(\theta)\,(u,u)]^*\, dW_t^\alpha \,| > \frac{r}{2\,\varepsilon}\, s_2(r) \ \} \ ,$$

the following estimate holds : On the *good* set $\Omega \setminus \Omega'_{\rm II}$

$$| - F_\alpha^\varepsilon(u,y,\varepsilon\, u) + \varepsilon\, H_\alpha(u,y,\varepsilon\, u)| \leq \tfrac{1}{2}\, r\, s_2(r)\, |u|^2$$

for any $u \in N_y^{r/\varepsilon} M_\alpha$. Using (13) and the estimate $|e^a - 1| \leq |a|\, e^{|a|}$ yields

$$\begin{aligned}|\langle \delta_{\rm II}, f\rangle| \ \ &\leq \ \ \tfrac{1}{2}\, s_1(r)\, r\, s_2(r)\, \|f\|_W \int_{M_\alpha} \int_{N_y M_\alpha} |u|^2\, W(|u|)\ \exp\{\tfrac{1}{2}\, r\, s_2(r)\, |u|^2\} \\ &\qquad\qquad \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\ p(y)\, \lambda_\alpha(dy)\ .\end{aligned}$$

Using the representation (9) and straightforward estimates for Gaussian integrals, yields

$$\begin{aligned}&\int_{N_y M_\alpha} |u|^2\, W(|u|)\ \exp\{\tfrac{1}{2}\, r\, s_2(r)\, |u|^2\}\ \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\ p(y)\, \lambda_\alpha(dy) \\ = \ &\int_{N_y M_\alpha} |u|^2\, W(|u|)\ \exp\{\tfrac{1}{2}\, r\, s_2(r)\, |u|^2\}\, \Gamma_\alpha(y,du)\ \bar{\mu}_\alpha(dy) \\ \leq \ & s_3^{k/2}(r)\ \exp\{\tfrac{1}{2}\, r\, s_2(r)\, s_3(r)\, |\chi_\alpha(y)|^2\}\ \bar{\mu}_\alpha(dy) \\ &\qquad \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} [\, s_3(r)\, |\chi_\alpha(y)| + \frac{\sqrt{s_3(r)}}{\sqrt{\lambda_{\rm min}}}\, |v|\,]^2\ W(s_3(r)\, |\chi_\alpha(y)| + \frac{\sqrt{s_3(r)}}{\sqrt{\lambda_{\rm min}}}\, |v|\,)\ \exp\{-\tfrac{1}{2}\, |v|^2\}\, dv\ ,\end{aligned}$$

where

$$s_3(r) = \frac{\lambda_{\rm min}}{\lambda_{\rm min} - r\, s_2(r)} \geq 1\ .$$

Notice that, using (6) yields

$$|\chi_\alpha(y)| \leq \frac{1}{\lambda_{\rm min}}\, |\zeta_\alpha(y)| \leq \frac{1}{\lambda_{\rm min}}\, S_1\ , \tag{15}$$

where the r.v.

$$S_1 = \sup_{\theta\in\Theta} \sup_{|u|=1} | \int_0^T [\dot{m}_t(\theta)\, u]^*\, dW_t^\alpha \,|$$

is $\mathbf{P}_\alpha^\varepsilon$–a.s. finite for any $\varepsilon > 0$, under Assumption B. Introducing the *bad* set

$$\Omega''_{\rm II} = \{ \sup_{\theta\in\Theta} \sup_{|u|=1} | \int_0^T [\dot{m}_t(\theta)\, u]^*\, dW_t^\alpha \,| > a\, \sqrt{\lambda_{\rm min}}\ \}\ ,$$

the following estimate holds : On the *good* set $\Omega \setminus (\Omega'_{\rm II} \cup \Omega''_{\rm II})$

$$\begin{aligned}&\frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} [\, s_3(r)\, |\chi_\alpha(y)| + \frac{\sqrt{s_3(r)}}{\sqrt{\lambda_{\rm min}}}\, |v|\,]^2\ W(s_3(r)\, |\chi_\alpha(y)| + \frac{\sqrt{s_3(r)}}{\sqrt{\lambda_{\rm min}}}\, |v|\,)\ \exp\{-\tfrac{1}{2}\, |v|^2\}\, dv \\ &\quad \leq c_W\, |\frac{2\, s_3(r)}{\sqrt{\lambda_{\rm min}}}|^2\ W(\frac{2\, s_3(r)}{\sqrt{\lambda_{\rm min}}})\ \ [\ a^2\, W(a) + \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} |v|^2\, W(|v|)\ \exp\{-\tfrac{1}{2}\, |v|^2\}\, dv\ ]\ ,\end{aligned}$$

hence

$$\frac{|\langle \delta_{\rm II}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} \leq \tfrac{1}{2}\, c_W\, s_1(r)\, s_2(r)\, s_3^{k/2+1}(r)\, |\frac{2\, s_3(r)}{\sqrt{\lambda_{\rm min}}}|^2\, W(\frac{2\, s_3(r)}{\sqrt{\lambda_{\rm min}}})\, \exp\{\tfrac{1}{2}\, s_2(r)\, \frac{s_3(r)}{\lambda_{\rm min}}\, r\, a^2\}$$
$$r\, \Big[\, a^2\, W(a) + \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} |v|^2\, W(|v|)\, \exp\{-\tfrac{1}{2}\, |v|^2\}\, dv\, \Big]\, \|f\|_W\ , \qquad (16)$$

which is finite under Assumption E.

▶ **Regular perturbation :** The next approximation is to take $\varepsilon = 0$ in the definition (12) of $\bar\rho_{\rm II}$, except for the indicator function, i.e. to define

$$\langle \bar\rho_{\rm III}, f\rangle = \int_{M_\alpha} f(y,y,u) \int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\,|u| < r)}\, \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\, p(y)\, \lambda_\alpha(dy)\ ,$$

and the corresponding error $\delta_{\rm III} = \bar\rho_{\rm II} - \bar\rho_{\rm III}$. It holds

$$|\langle \delta_{\rm III}, f\rangle| \leq \int_{M_\alpha} \int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\,|u| < r)} |\omega(y,\varepsilon\, u, u)|\, \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\, p(y)\, \lambda_\alpha(dy)\ ,$$

where

$$\begin{aligned} \omega(y,z,u) &= f(y+z,y,u)\, \frac{p(y+z)}{p(y)}\, J_\alpha(y,z) - f(y,y,u) \\ &= [f(y+z,y,u) - f(y,y,u)] + f(y+z,y,u)\, \frac{p(y+z)}{p(y)}\, [J_\alpha(y,z) - 1] \\ &\qquad + f(y+z,y,u)\, [\frac{p(y+z)}{p(y)} - 1]\ . \end{aligned}$$

For any $(y,z) \in N^r M_\alpha$, it holds

$$|\omega(y,z,u)| \leq [\, r\, \|f\|_{{\rm L}(W)} + r\, s_4(r)\, \|f\|_W + \omega(r)\, \|f\|_W\,]\, W(|u|)\ ,$$

where

$$s_4(r) = \sup_{(y,z)\in N^r M_\alpha} \frac{p(y+z)}{p(y)}\, \frac{|J_\alpha(y,z) - 1|}{|z|} \qquad \text{and} \qquad \omega(r) = \sup_{(y,z)\in N^r M_\alpha} |\, \frac{p(y+z)}{p(y)} - 1|\ .$$

Notice that $s_4(r)$ is finite for small $r > 0$, and $\omega(r)$ can be made arbitrarily small with r, since p is continuous, J_α is Lipschitz continuous, see e.g. Berger and Gostiaux [1, Corollary 6.8.11], and $N^r M_\alpha$ is compact. Using the representation (9) and straightforward estimates for Gaussian integrals, yields

$$\begin{aligned} &\int_{N_y M_\alpha} W(|u|)\, \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\, p(y)\, \lambda_\alpha(dy) \\ &= \int_{N_y M_\alpha} W(|u|)\, \Gamma_\alpha(y,du)\, \bar\mu_\alpha(dy) \\ &\leq \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} W(|\chi_\alpha(y)| + \frac{1}{\sqrt{\lambda_{\rm min}}}\, |v|\,)\, \exp\{-\tfrac{1}{2}\, |v|^2\}\, dv\, \bar\mu_\alpha(dy)\ . \end{aligned}$$

Introducing the *bad* set

$$\Omega_{\rm III} = \{\, \sup_{\theta\in\Theta} \sup_{|u|=1} |\int_0^T [\dot m_t(\theta)\, u]^*\, dW_t^\alpha\, | > c\, \sqrt{\lambda_{\rm min}}\ \}\ ,$$

and using (15), the following estimate holds : On the *good* set $\Omega \setminus \Omega_{\rm III}$

$$\frac{|\langle \delta_{\rm III}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} \leq c_W\, W(\frac{2}{\sqrt{\lambda_{\rm min}}})\, [\, r + r\, s_4(r) + \omega(r)\,]$$
$$\Big[\, W(c) + \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} W(|v|)\, \exp\{-\tfrac{1}{2}\, |v|^2\}\, dv\, \Big]\, \|f\|_{{\rm BL}(W)}\ , \qquad (17)$$

which is finite under Assumption E.

▶ **Integration over the unbounded normal bundle space :** The last step is to replace $\bar{\rho}_{\rm III}$ with $\bar{\rho}_\alpha$, and to define the corresponding error $\delta_{\rm IV} = \bar{\rho}_{\rm III} - \bar{\rho}_\alpha$. It holds

$$|\langle \delta_{\rm IV}, f\rangle| \leq \|f\|_W \int_{M_\alpha} \int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\,|u| > r)}\, W(|u|)\; \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\, p(y)\, \lambda_\alpha(dy)\;.$$

Using the representation (9) and straightforward estimates for Gaussian integrals, yields

$$\begin{aligned}
&\int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\,|u| > r)}\, W(|u|)\; \exp\{-Q_\alpha(y,u)\}\, \lambda_{\alpha,y}(du)\, p(y)\, \lambda_\alpha(dy) \\
&\quad = \int_{N_y M_\alpha} \mathbf{1}_{(\varepsilon\,|u| > r)}\, W(|u|)\, \Gamma_\alpha(y,du)\; \bar{\mu}_\alpha(dy) \\
&\quad \leq \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} \mathbf{1}_{(|v| > \frac{r}{2\,\varepsilon}\,\lambda_{\min})}\, W(|v|)\; \exp\{-\tfrac{1}{2}\,|v|^2\}\, dv\; \bar{\mu}_\alpha(dy)\;,
\end{aligned}$$

provided

$$|\chi_\alpha(y)| \leq \frac{r}{2\,\varepsilon}\;,$$

which, using (15), is guaranteed if

$$S_1 \leq \frac{r}{2\,\varepsilon}\,\lambda_{\min}\;.$$

Introducing the *bad* set

$$\Omega_{\rm IV} = \{\, \sup_{\theta\in\Theta}\; \sup_{|u|=1}\; |\int_0^T [\dot{m}_t(\theta)\,u]^*\, dW_t^\alpha\,| > \frac{r}{2\,\varepsilon}\,\lambda_{\min}\,\}\;,$$

the following estimate holds : On the *good* set $\Omega \setminus \Omega_{\rm IV}$,

$$\frac{|\langle \delta_{\rm IV}, f\rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1}\rangle} \leq \|f\|_W\, \frac{1}{(2\,\pi)^{k/2}} \int_{\mathbb{R}^k} \mathbf{1}_{(|v| > \frac{r}{2\,\varepsilon}\,\lambda_{\min})}\, W(|v|)\; \exp\{-\tfrac{1}{2}\,|v|^2\}\, dv\;, \tag{18}$$

which is finite under Assumption E.

▶ **Final step :** Notice that

$$\begin{aligned}
\|\rho^\varepsilon - \rho_\alpha\|^*_{{\rm BL}(W)} &= \sup_{\|f\|_{{\rm BL}(W)}=1} |\langle \rho^\varepsilon, f\rangle - \langle \rho_\alpha, f\rangle| \\
&\leq \sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \bar{\rho}^\varepsilon, f\rangle - \langle \bar{\rho}_\alpha, f\rangle|}{\langle \bar{\rho}_\alpha, \mathbf{1}\rangle} + \sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \bar{\rho}^\varepsilon, f\rangle|}{\langle \bar{\rho}^\varepsilon, \mathbf{1}\rangle}\, \frac{|\langle \bar{\rho}^\varepsilon, \mathbf{1}\rangle - \langle \bar{\rho}_\alpha, \mathbf{1}\rangle|}{\langle \bar{\rho}_\alpha, \mathbf{1}\rangle} \\
&\leq 2 \sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \bar{\rho}^\varepsilon, f\rangle - \langle \bar{\rho}_\alpha, f\rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1}\rangle}\;,
\end{aligned}$$

since $\langle \bar{\rho}_\alpha, \mathbf{1}\rangle = \langle \bar{\mu}_\alpha, \mathbf{1}\rangle$, hence for any $\eta > 0$

$$\begin{aligned}
\mathbf{P}^\varepsilon_\alpha[\,\|\rho^\varepsilon - \rho_\alpha\|^*_{{\rm BL}(W)} > \eta] &\leq \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \bar{\rho}^\varepsilon, f\rangle - \langle \bar{\rho}_\alpha, f\rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1}\rangle} > \tfrac{1}{2}\,\eta] \\
&\leq \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \delta_{\rm I}, f\rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta] + \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \delta_{\rm II}, f\rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta] \\
&\qquad + \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \delta_{\rm III}, f\rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta] + \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{{\rm BL}(W)}=1} \frac{|\langle \delta_{\rm IV}, f\rangle|}{\langle \bar{\mu}_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta]\;.
\end{aligned}$$

We then intersect each of the above events by their corresponding *good* sets, i.e.

$$
\begin{aligned}
\mathbf{P}^\varepsilon_\alpha[\,\|\rho^\varepsilon - \rho_\alpha\|^*_{\mathrm{BL}(W)} > \eta] \quad \leq \quad & \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{I}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus\Omega_{\mathrm{I}}] + \mathbf{P}^\varepsilon_\alpha[\Omega_{\mathrm{I}}] \\
& + \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{II}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus(\Omega'_{\mathrm{II}}\cup\Omega''_{\mathrm{II}})] + \mathbf{P}^\varepsilon_\alpha[\Omega'_{\mathrm{II}}] + \mathbf{P}^\varepsilon_\alpha[\Omega''_{\mathrm{II}}] \\
& + \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{III}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus\Omega_{\mathrm{III}}] + \mathbf{P}^\varepsilon_\alpha[\Omega_{\mathrm{III}}] \\
& + \mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{IV}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus\Omega_{\mathrm{IV}}] + \mathbf{P}^\varepsilon_\alpha[\Omega_{\mathrm{IV}}]\ .
\end{aligned}
$$

From estimate (11) it follows that

$$
\mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{I}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus\Omega_{\mathrm{I}}] = 0\ ,
$$

provided $\dfrac{r^2}{\varepsilon}$ is large enough, and ε is small enough. From estimate (16) it follows that

$$
\mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{II}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus(\Omega'_{\mathrm{II}}\cup\Omega''_{\mathrm{II}})] = 0\ ,
$$

provided r and $r\,a^2\,W(a)$ are small enough. From estimate (17) it follows that

$$
\mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{III}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus\Omega_{\mathrm{III}}] = 0\ ,
$$

provided r, $r\,W(c)$ and $\omega(r)\,W(c)$ are small enough. Finally, from estimate (18) it follows that

$$
\mathbf{P}^\varepsilon_\alpha[\sup_{\|f\|_{\mathrm{BL}(W)}=1} \frac{|\langle \delta_{\mathrm{IV}}, f\rangle|}{\langle \bar\mu_\alpha, \mathbf{1}\rangle} > \tfrac{1}{8}\,\eta\,,\,\Omega\setminus\Omega_{\mathrm{IV}}] = 0\ ,
$$

provided $\dfrac{r}{\varepsilon}$ is large enough. On the other hand, notice that

$$
\Omega_{\mathrm{I}} \subset \{\ \sup_{\theta\in\Theta} |\int_0^T [m_t(\theta) - m_t(\alpha)]^*\, dW^\alpha_t\,| > \frac{r^2}{8\,\varepsilon}\,\lambda_{\min}\ \}\ ,
$$

for r small enough, hence $\mathbf{P}^\varepsilon_\alpha[\Omega_{\mathrm{I}}]$ is small when $\dfrac{r^2}{\varepsilon}$ is large. Notice also that

$$
\Omega'_{\mathrm{II}} \subset \{\ \sup_{\theta\in\Theta}\ \sup_{|u|=1} |\int_0^T [\ddot m_t(\theta)\,(u,u)]^*\, dW^\alpha_t\,| > \frac{r}{\varepsilon}\,\sigma_1\,\sigma_2\ \}\ ,
$$

hence $\mathbf{P}^\varepsilon_\alpha[\Omega'_{\mathrm{II}}]$ is small when $\dfrac{r}{\varepsilon}$ is large. Similarly, $\mathbf{P}^\varepsilon_\alpha[\Omega''_{\mathrm{II}}]$, $\mathbf{P}^\varepsilon_\alpha[\Omega_{\mathrm{III}}]$ and $\mathbf{P}^\varepsilon_\alpha[\Omega_{\mathrm{IV}}]$ are small when a, c and $\dfrac{r}{\varepsilon}$ are large, respectively. Therefore, taking $r = r(\varepsilon)$ such that

$$
\lim_{\varepsilon\downarrow 0} r(\varepsilon) = 0 \qquad \text{and} \qquad \lim_{\varepsilon\downarrow 0} \frac{r^2(\varepsilon)}{\varepsilon} = \infty\ ,
$$

which ensures that $\displaystyle\lim_{\varepsilon\downarrow 0} \frac{r(\varepsilon)}{\varepsilon} = \infty$, taking $a = a(\varepsilon)$ such that

$$
\lim_{\varepsilon\downarrow 0} a(\varepsilon) = \infty \qquad \text{and} \qquad \lim_{\varepsilon\downarrow 0} r(\varepsilon)\,a^2(\varepsilon)\,W(a(\varepsilon)) = 0\ , \tag{19}
$$

which ensures that $\displaystyle\lim_{\varepsilon\downarrow 0} r(\varepsilon)\,a^2(\varepsilon) = 0$, and taking $c = c(\varepsilon)$ such that

$$
\lim_{\varepsilon\downarrow 0} c(\varepsilon) = \infty\ , \qquad \lim_{\varepsilon\downarrow 0} r(\varepsilon)\,W(c(\varepsilon)) = 0 \qquad \text{and} \qquad \lim_{\varepsilon\downarrow 0} \omega(r(\varepsilon))\,W(c(\varepsilon)) = 0\ , \tag{20}
$$

is sufficient to guarantee that all probabilities of bad events go to zero as $\varepsilon\downarrow 0$.

▶ **Exponential rate :** It follows from Marcus and Shepp [18], see also Ledoux and Talagrand [16, Corollary 3.2], and from identities (1), (2) and (3), that

$$\begin{aligned}
\lim_{\lambda\uparrow\infty} \frac{1}{\lambda^2} \log \mathbf{P}^\varepsilon_\alpha[\sup_{\theta\in\Theta} | \int_0^T [m_t(\theta) - m_t(\alpha)]^* \, dW^\alpha_t \, | > \lambda \,] &= -\frac{1}{2\,\sigma_0^2} \;, \\
\lim_{\lambda\uparrow\infty} \frac{1}{\lambda^2} \log \mathbf{P}^\varepsilon_\alpha[\sup_{\theta\in\Theta} \sup_{|u|=1} | \int_0^T [\dot m_t(\theta)\, u]^* \, dW^\alpha_t \, | > \lambda \,] &= -\frac{1}{2\,\sigma_1^2} \;, \\
\lim_{\lambda\uparrow\infty} \frac{1}{\lambda^2} \log \mathbf{P}^\varepsilon_\alpha[\sup_{\theta\in\Theta} \sup_{|u|=1} | \int_0^T [\ddot m_t(\theta)\,(u,u)]^* \, dW^\alpha_t \, | > \lambda \,] &= -\frac{1}{2\,\sigma_2^2} \;.
\end{aligned}$$

Therefore

$$\lim_{\varepsilon\downarrow 0} \frac{1}{\lambda^2(\varepsilon)} \log \mathbf{P}^\varepsilon_\alpha[\, \|\rho^\varepsilon - \rho_\alpha\|^*_{\mathrm{BL}(W)} > \eta] < 0 \;,$$

where $\lambda^2(\varepsilon)$ denotes the slowest of the exponential rates $\dfrac{r^4(\varepsilon)}{\varepsilon^2}$, $a^2(\varepsilon)$ and $c^2(\varepsilon)$, subject to (19) and (20). If $W \equiv 1$, then (19) and (20) reduce to

$$\lim_{\varepsilon\downarrow 0} r(\varepsilon)\, a^2(\varepsilon) = 0 \;.$$

With the choice $r(\varepsilon) \sim \varepsilon^{2/5}$, the rate is $\lambda^2(\varepsilon) \sim \varepsilon^{-2/5}\,\delta(\varepsilon)$, where $\delta(\varepsilon)$ goes to zero arbitrarily slowly as $\varepsilon \downarrow 0$. If the enveloppe W is of the form $W(r) = \max(1, r^q)$ for some $q > 0$, and if in addition the prior density p is Lipschitz continuous on M_α, then $\omega(r(\varepsilon)) \sim r(\varepsilon)$, and (19) and (20) reduce to

$$\lim_{\varepsilon\downarrow 0} r(\varepsilon)\, a^{q+2}(\varepsilon) = 0 \qquad \text{and} \qquad \lim_{\varepsilon\downarrow 0} r(\varepsilon)\, c^q(\varepsilon) = 0 \;.$$

With the choice $r(\varepsilon) \sim \varepsilon^{(q+2)/(2q+5)}$, the rate is $\lambda^2(\varepsilon) \sim \varepsilon^{-2/(2q+5)}\,\delta(\varepsilon)$, where $\delta(\varepsilon)$ goes to zero arbitrarily slowly as $\varepsilon \downarrow 0$.

Contents

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399